



Université catholique de Louvain

Analyse numérique:

Approximation, interpolation, intégration.

INMA2171

2010-2011

Alphonse Magnus,
Institut de Mathématique Pure et Appliquée,
Université Catholique de Louvain,
Chemin du Cyclotron, 2,
B-1348 Louvain-la-Neuve
(Belgium)

(0)(10)473157 , alphonse.magnus@uclouvain.be ,
<http://perso.uclouvain.be/alphonse.magnus/>



Groucho Marx dans *Animal Crackers*

Pourquoi les chiffres fascinent-ils tant de simples d'esprit et les impatients toujours friands de références et de certitudes? *Un chiffre ne se discute pas*, en quelque sorte par définition; il y a bien une virilité imbécile du chiffre entêté et toujours prêt à s'abriter derrière une espèce d'immunité scientifique.

Gilles Châtelet, *Vivre et penser comme des porcs*, Exils 1998 = Folio Actuel 73, 1999.

Je suis fatigué des chiffres
mais amoureux de leur
monde truqué
John Maeda

**L'abus de
nombres
 nuit à
la santé**

*A Multi Millionaire
is Talking Amicably
to His Man Servant*

*Paxton, look how, when I
zip my thumb across this
cheque book stub, the
numbers seem to
move...*



John Glashan, *The jokes of John Glashan*, Private Eye, London, 1975.



Don Martin, *MAD Magazine* 111, June 1967. Voir suite p. 183

ELEMENTARY TRANSCENDENTAL FUNCTIONS
RADIX TABLE OF NATURAL LOGARITHMS Table 4.

x	n	$\ln (1+x10^{-n})$	$-\ln (1-x10^{-n})$
1	5	0.0000 99999 50000 33333 08334	0.0001 00000 50000 33333 58334
2	5	0.00001 99998 00002 66662 66673	0.00002 00002 00002 66670 66673
3	5	0.00002 99995 50008 99979 75049	0.00003 00004 50009 00020 25000
4	5	0.00003 99992 00021 33269 33538	0.00004 00008 00021 33397 33538
5	5	0.00004 99987 50041 66510 42292	0.00005 00012 50041 66822 92292
6	5	0.00005 99982 00071 99676 01555	0.00006 00018 00072 00324 01555



Table des matières.

Préface.	8
.....	9
Analyse numérique et théorie de l'approximation.	12
1. Qu'est ce que l'analyse numérique?	12
1.1. Analyse numérique et analyse	12
1.2. Analyse numérique et calcul	12
2. Théorie de l'approximation.	13
2.1. Les trois niveaux d'une théorie de l'approximation.	13
3. Quelques approximations de fonctions utilisées dans les calculatrices et les ordinateurs	
14	
3.1. Calculatrices scientifiques: le système CORDIC	14
3.2. Approximations polynomiales et rationnelles	15
3.3. AGM, etc	16
3.4. Approximations et nombres irrationnels	16
3.5. Approximations les plus simples: bien commencer	18
CHAPITRE 1. Théorèmes généraux d'existence et d'unicité de meilleure approximation.	
19	
1. Distances et normes.	19
1.1.	19
1.2. Exercices et exemples	19
1.3. Remarques	19
1.4. Normes	20
1.5. Exemples, exercices	20
1.6. Exercices, exemples	21
1.7. Formes et applications linéaires continues sur des espaces vectoriels normés de fonctions	
21	
2. Existence d'une meilleure approximation.	22
2.1. Théorème d'existence de meilleure approximation dans un sous-espace de dimension finie	
22	
2.2. Contre-exemple	23
2.3. Remarque	23
3. Unicité de la meilleure approximation.	23
3.1. Définition. Convexité	23
3.2. Proposition	23
3.3. Définition	24
3.4. Une condition suffisante d'unicité. Théorème	24

3.5. Exercice	24
4. Continuité du projecteur de meilleure approximation.	24
4.1. Théorème de continuité	24
4.2. Forte unicité	25
5. Dualité.	25
6. Exemples et exercices.	25
6.1.	25
6.2. Moyenne et médiane	26
6.3. Principaux sous-espaces de fonctions utilisés en approximation	26
6.4. Centre et rayon de Tchebycheff d'une partie P de X	27
6.5. Largeurs de Kolmogorov	27
6.6. Coapproximation	27
CHAPITRE 2. Approximation au sens de Tchebycheff.	28
1. Théorème d'équioscillation de Tchebycheff.	28
1.1. Théorème d'équioscillation de Tchebycheff (1853)	28
1.2. Preuve de la condition nécessaire: \hat{p} optimal dans $\mathcal{P}_n \Rightarrow (3)$	29
1.3. Preuve de la condition suffisante $(3) \Rightarrow \hat{p}$ optimal dans \mathcal{P}_n	30
1.4. Exemple	30
1.5. Théorème d'unicité de la meilleure approximation polynomiale au sens de Tchebycheff	
30	
2. Propriétés de la meilleure approximation.	31
2.1. Symétrie. Théorème	31
2.2. Théorème (de La Vallée Poussin)	31
2.3. Unicité forte	32
2.4. Signes alternés	32
2.5. Algorithme d'échange	33
2.6. Meilleure approximation au sens $\ \cdot \ _\infty$ sur un compact quelconque	34
3. Polynômes de Tchebycheff.	35
3.1. Meilleure approximation d'un polynôme de degré n dans \mathcal{P}_{n-1}	35
3.2. Définition	35
3.3. Premières propriétés	35
3.4. Premiers échantillons	37
3.5. Exercice	39
3.6. Relation de récurrence	39
3.7. Polynôme de moindre norme sur un intervalle, sous contrainte $p(0) = 1$	40
3.8. Meilleure approximation d'un polynôme de degré n dans \mathcal{P}_{n-2}	40
3.9. Meilleure approximation de fonction rationnelle	40
3.10. Fonctions rationnelles de moindre et plus grande déviation	40
3.11. Propriétés extrémales des polynômes de Tchebycheff	41
3.12. Autres propriétés des T_n	44
3.13. Equation différentielle linéaire	47
3.14. Proposition	47
3.15. Polynômes de Tchebycheff et problèmes de Sturm-Liouville	48
3.16. Coefficients, dérivées et primitives	49
3.17. Primitives itérées de $\frac{T_n(x)}{\sqrt{1-x^2}}$ et formule de Rodrigues	50

3.18.	Orthogonalité et base duale de \mathcal{P}_N^*	51
4.	Bonne approximation; séries de polynômes de Tchebycheff, relation avec séries de Fourier. 54	
4.1.	Bonne base et bonne approximation	54
4.2.	Cascade de meilleures approximations et développements dans la base des polynômes de Tcheb 55	
4.3.	Série de Fourier d'une fonction continue périodique	56
4.4.	Séries de polynômes de Tchebycheff	57
4.5.	Vitesse de décroissance des coefficients et bornes de norme de fonction d'erreur	60
4.6.	Théorème de Weierstrass	62
4.7.	Calcul des coefficients de Tchebycheff et autres algorithmes	63
4.8.	Algorithmes en représentation de Tchebycheff	66
5.	Approximation par fonction rationnelle.	67
6.	Lecture. Tchebycheff et de La Vallée Poussin	68
CHAPITRE 3. Approximation en moyenne quadratique.		73
1.	Produit scalaire, orthogonalité, espace préhilbertien.	73
1.1.	Produits scalaires sur \mathcal{P}_n	73
1.2.	Espace préhilbertien	76
2.	Meilleure approximation dans un espace préhilbertien.	76
2.1.	Base d'un espace de dimension finie, matrice de Gram	77
2.2.	Meilleure approximation = projection orthogonale	78
2.3.	Méthode d'orthogonalisation de Gram-Schmidt	80
2.4.	Hauteurs, volumes et déterminants de Gram	81
2.5.	Factorisation de Cholesky	82
3.	Polynômes orthogonaux.	84
3.1.	Construction d'une base orthogonale de \mathcal{P}_n	84
3.2.	Relation de récurrence	88
3.3.	Quelques algorithmes utilisant la récurrence	89
3.4.	Zéros des polynômes orthogonaux	93
3.5.	Zéros de polynômes orthogonaux et valeurs propres de matrices tridiagonales symétriques 94	
3.6.	Formules d'intégration de Gauss. Première approche	96
3.7.	Formule d'intégration de Gauss et fractions continues	98
3.8.	Formule de Christoffel-Darboux	100
3.9.	Orthogonalité et opérateurs (formellement) hermitiens	101
3.10.	Polynômes orthogonaux classiques	102
3.11.	Usages et variétés de polynômes orthogonaux	106
3.12.	Harmoniques sphériques et fonctions de Legendre	109
3.13.	Polynômes d'Hermite et mécanique quantique	112
3.14.	Orthogonalité et équioscillation	113
4.	Moindres carrés, régression.	114
5.	Approximation en norme $\ \cdot \ _1$	118
6.	Séries de Fourier en analyse numérique.	118
6.1.	Comportement des coefficients	121
6.2.	Transformée de Fourier discrète	122
6.3.	Transformée de Fourier rapide (Fast Fourier Transform FFT)	124

6.4.	Analyse en ondelettes	126
7.	Convergence, espace de Hilbert	129
7.1.	Suites totales et maximales	129
7.2.	Théorème.	130
7.3.	Exemples de suites totales dans $\mathcal{C}[a, b]$ et $L^2([a, b], \mu)$.	131
7.4.	Théorème d'approximation de Weierstrass.	131
7.5.	Théorème de Stone-Weierstrass	135
7.6.	Noyaux reproduisants, polynômes noyaux, représentants de Riesz	137
7.7.	Intervalles non bornés, problème des moments.	139
7.8.	Arcs de Bézier en typographie informatique	141
CHAPITRE 4. Interpolation et applications.		143
1.	Interpolation.	143
1.1.	Interpolation polynomiale classique	144
1.2.	Interpolation: cadre général	146
1.3.	Interpolation polynomiale classique en formulation de Newton, différences divisées.	150
1.4.	Extrapolation à la limite de Richardson	152
1.5.	Formulation de Newton en général	153
1.6.	Différences divisées et dérivées	155
1.7.	Reste de l'interpolation polynomiale classique	157
1.8.	Interpolation d'Hermite-Fejér	158
2.	Formules d'intégration basées sur l'interpolation.	159
2.1.	Reste de la formule de quadrature de Gauss	159
2.2.	Formules de quadratures de Gauss avec points imposés	160
2.3.	Points de Tchebycheff: règle de Clenshaw-Curtis	161
2.4.	Règles adaptatives d'intégration	161
3.	Représentation du reste: théorème de Peano.	162
3.1.	Théorème (Peano)	162
CHAPITRE 5. Différences finies.		165
1.	Les opérateurs du calcul aux différences.	165
2.	Interpolation.	168
3.	Dérivation.	169
4.	Intégration.	170
4.1.	Formules de Newton-Cotes	170
5.	Noyaux de Peano de règles d'intégration.	171
5.1.	Formule du trapèze	171
5.2.	Formule de Simpson	172
5.3.	Noyaux de Peano de formules composées	172
5.4.	La formule de Simpson avant Simpson	173
6.	Formule d'Euler-Maclaurin.	174
6.1.	Identités des nombres et polynômes de Bernoulli	176
6.2.	Schéma d'intégration de Romberg.	179
6.3.	Formule d'Euler-Maclaurin en tant que formule sommatoire	180
CHAPITRE 6. Récapitulation.		184

CHAPITRE 7. Appendices: alphabets grec, cyrillique, petit dico, index.	185
1. Alphabets	185
2. Petit dico mathematical English \rightarrow français mathématique.	186
Index	187

Préface.

L'analyse numérique a longtemps été incorporée au cours d'analyse générale, dont elle représentait le versant appliqué et constructif.

Le fort développement des moyens de calcul automatique a rendu nécessaire l'apparition d'un enseignement spécifique. La discipline put se développer sous l'impulsion de mathématiciens avisés, tels P. Henrici [**Hen**] et E. Stiefel [**Sti**].

Le présent cours fut créé par *Jean Meinguet*, Professeur à l'Université. On trouvera ici l'essentiel de la partie "approximation, interpolation, intégration" de son enseignement. D'autres cours reprennent les thèmes de résolution numériques des équations (y compris différentielles et fonctionnelles), d'algèbre linéaire numérique (théorie des matrices) et d'algorithmique numérique.

Le Professeur Meinguet est également à l'origine de l'enseignement de la programmation et de l'informatique dans notre université, mais cela est une autre histoire. . .

Les grands principes de la théorie de l'approximation sont d'abord déduits de concepts d'analyse fonctionnelle (chap. 1). Ces résultats sont alors appliqués à des situations plus concrètes: on examine en détail l'approximation par des polynômes et par des polynômes trigonométriques, selon la norme du maximum (chap. 2) et en moyenne quadratique (chap. 3). Avec l'interpolation (chap. 4) et le calcul aux différences finies (chap. 5), on dispose des outils permettant de traiter tous les problèmes de l'analyse numérique classique, en particulier l'intégration numérique.

- Quelques ouvrages fréquemment cités ici (par une mention du type [xxx]), normalement disponibles en bibliothèques de MATH, PHYS et BSE:
- [Abr] M. Abramowitz, I.A. Stegun, ed., *Handbook of Mathematical Functions*, Nat. Bureau of Standards, Washington, 1964 = Dover, New York, 1965 etc. <http://members.fortunecity.com/aands/>, <http://www.convertit.com/Go/ConvertIt/Reference/AMS55.ASP>
<http://www.math.sfu.ca/~cbm/aands/> Abramowitz and Stegun: Handbook of Mathematical Functions
 An electronic copy of the tenth printing of this famous reference. Voir aussi David Alan Grier: "Irene Stegun, the Handbook of Mathematical Functions, and the Lingering Influence of the New Deal," *The American Mathematical Monthly*, Vol. **113**, No. 7, August-September 2006, pp. 585-597.
- [Ach] N.I. Achieser, *Theory of Approximation*, F. Ungar, 1956 =Dover, 1992.
- [Atk] K. Atkinson, W. Han, *Theoretical Numerical Analysis, A Functional Analysis Framework*, Springer texts in Applied Mathematics **39**, Springer 2001; *Elementary Numerical Analysis*, 3rd edition, John Wiley, 2003. Notes et programmes dans <http://www.math.uiowa.edu/~atkinson/>
- [Che] E.W. Cheney, *Introduction to Approximation Theory*, McGraw-Hill, 1966.
- [CheK] E.W. Cheney, D. Kincaid, *Numerical Mathematics and Computing*, Brooks/Cole, 4th ed., 1999.
- [CheL] E.W. Cheney, W. Light, *A Course in Approximation Theory*, Brooks/Cole, 1999.
- [Chi] T.S. Chihara, *An Introduction to Orthogonal Polynomials*, Gordon & Breach, New York, 1978.
- [Clen] C.W. Clenshaw, *Math. Tables 5 Chebyshev Series for Mathematical Functions*, Nat. Physical Laboratory, London: Her Majesty's Stationery Office, 1962.
- [Dav] P.J. Davis, *Interpolation and Approximation*, Blaisdell, Waltham, 1963 = Dover, New York, 1975.
- [DaR] P.J. Davis, Ph. Rabinowitz, *Methods of Numerical Integration*, 2^{ème} édition, Academic Press, New York, 1984.
- [deB] C. de Boor, *Numerical Functional Analysis*, notes du cours CS717 de l'Université du Wisconsin <http://www.cs.wisc.edu/~deboor/717/notes.html>
- [DeVLor] R.A. DeVore, G.G. Lorentz, *Constructive Approximation*, Springer, Berlin, 1993.
- [Erd] A. Erdélyi, W. Magnus, F. Oberhettinger, F.G. Tricomi, *Higher Transcendental Functions*, 3 vol., *Tables of Integral Transforms*, 2 vol., (*The Bateman Manuscript Project*), McGraw-Hill, New York, 1953-1955.
- [FoxP] L. Fox, I.B. Parker, *Chebyshev Polynomials in Numerical Analysis*, Oxford U.P., 1968.
- [GasW] C. Gasquet, P. Witomski, *Analyse de Fourier et applications. Filtrage, calcul numérique, ondelettes*, Masson, Paris, 1990.
- [Gau] W. Gautschi, *Numerical Analysis. An Introduction*, Birkhäuser, 1997.
- [Gau2] W. Gautschi, *Orthogonal Polynomials: Computation and Approximation*, Oxford U. Press, 2004.
- [Hai] E. Hairer, *Introduction à l'analyse numérique*, cours Université de Genève, 1993, cf "Polycopiés" dans <http://www.unige.ch/math/folks/hairer/>
- [Hammer] G. Hämmerlin, K.H. Hoffmann, *Numerische Mathematik*, Springer-Verlag, 1989 = *Numerical Mathematics*, Springer-Verlag, New York, 1991.
- [Ham] R.W. Hamming, *Numerical Methods for Scientists and Engineers*, 2nd ed., McGraw-Hill, New York, 1973 = Dover
- [Has] C. Hastings, Jr., *Approximations for Digital Computers*, Princeton U.P., 1955.
- [Hen] P. Henrici, *Elements of Numerical Analysis*, Wiley, New York, 1964.
- [Hen82] P. Henrici, *Essentials of Numerical Analysis*, Wiley, New York, 1982.
- [Kah] D. Kahaner, C. Moler, S. Nash, *Numerical Methods and Software*, Prentice-Hall, 1989.
- [Kun] J. Kuntzmann, *Méthodes numériques*, Hermann, Paris, 1969.
- [Lanc] C. Lanczos, *Applied Analysis*, Prentice Hall, Englewood Cliffs, 1956 = Dover
- [MilneT] L. M. Milne Thomson, *The Calculus Of Finite Differences*, Macmillan And Company., Limited, 1933, <http://www.archive.org/details/calculusoffinite032017mbp>
- [Mei] G. Meinardus, *Approximation von Funktionen und ihre numerische Behandlung*, Springer, Berlin, 1964 = *Approximation of Functions: Theory and Numerical Methods*, Springer, 1967.
- [Mha] Mhaskar, Hrushikesh N.; Pai, Devidas V., *Fundamentals of approximation theory*, Alpha Science International, 2000.
- [Nat] I.P. Natanson, *Constructive Theory of Functions*, Moscou-Leningrad 1949 = Translation Series, U.S. Atomic Energy Commission AET-tr-4503.
- [Nurn] G. Nürnberger, *Approximation by spline functions*, Springer-Verlag. Berlin, 1989.
- [Pas] S. Paszkowski, *Polynômes et séries de Tchebichev*, Report ANO 140, Univ. Lille1, Juillet 1984.
- [Pow] M.J.D. Powell, *Approximation Theory and Methods*, Cambridge U.P., Cambridge, 1981.
- [RalR] A. Ralston, P. Rabinowitz, *A First Course in Numerical Analysis*, 2nd ed., McGraw-Hill, 1978.

- [Rap] J. Rappaz, M. Picasso, *Introduction à l'analyse numérique*, PPUR (Presses Polytechniques et Universitaires Romandes), Lausanne, 1998, <http://dmawww.epfl.ch/rappaz.mosaic>
- [Riv1] T.J. Rivlin, *An Introduction to the Approximation of Functions*, Blaisdell, Waltham, 1969 = Dover, New York, 1981.
- [Riv2] T.J. Rivlin, *The Chebyshev Polynomials, From Approximation Theory to Algebra and Number Theory*, Wiley, New York, 2^{ème} édition, 1990.
- [Sch] F. Scheid, *Schaum's outline of theory and problems of numerical analysis*, 2nd ed., McGraw-Hill 1988.
- [Sti] E. Stiefel, *Introduction à la mathématique numérique*, Dunod, Paris 1967 = *Einführung in die numerische Mathematik*, Teubner, Stuttgart, 1961 = *An Introduction to Numerical Mathematics*, Academic Press, New York 1963.
- [Sze] G. Szegő, *Orthogonal Polynomials*, 4th ed., Amer. Math. Soc. , Colloquium Publications, vol. 23, Providence, 1975.
- [Tche] P.L. Tchebychef, *Œuvres, publiées par les soins de MM. A. Markoff et N. Sonin*, 2vol., Chelsea, New York. <http://www.archive.org/details/oeuvresdepltche01chebrich>
<http://www.archive.org/details/oeuvresdepltche02chebrich>
- [dLVP] C.J. de La Vallée Poussin, *Leçons sur l'approximation des fonctions d'une variable réelle*, Gauthier-Villars, Paris, 2^{ème} édition, 1919 = Chelsea, New York, 1970.
<http://www.archive.org/details/LeconsSurApproximadesFonctions>
- [dLVP2] C.J. de La Vallée Poussin, *Œuvres*, rassemblées dans la bibliothèque MATH. 3 volumes parus, Académie royale de Belgique & Circolo Matematico di Palermo.

Périodiques.

<http://www.ams.org/mathweb/mi-journals.html>

ACM¹ Transactions on Mathematical Software, <http://math.nist.gov/toms/>, Advances in Computational Mathematics, Annals of Numerical Mathematics, Applied Numerical Mathematics, Approximation Theory and its Applications, Calcolo, Constructive Approximation, Electronic Transactions on Numerical Analysis <http://etna.mcs.kent.edu/html/>, Journal of Approximation Theory, Journal of Computational and Applied Mathematics <http://www.elsevier.nl/locate/cam>, Mathematics of Computation, Numerical Algorithms, Numerische Mathematik, SIAM² Journal of Numerical Analysis.

Ressources réseau.

- e-livres Livres de mathématiques numérisés <http://edmond.orignac.pagesperso-orange.fr/math-refs.html>
- Usenet <news:sci.math.num-analysis>
- FAQ FAQ: Numerical Analysis and Associated Fields Resource Guide, by Steve Sullivan (Mathcom, Inc.), voir "Tech Info" dans <http://www.mathcom.com/>
- Lille École d'ingénieurs de Lille <http://www.eudil.fr/>
- Conc Numerical Analysis at Concordia. <http://indy.cs.concordia.ca/na/>
- ATNet Approximation Theory net. Contient des notes de cours (voir à "Classroom notes")
<http://www.mi.uni-erlangen.de/at-net>
- Matlab, Java <http://www.mathworks.com>, <http://www.mathtools.net>
- Matlab, Atkinson <ftp://ftp.math.uiowa.edu/pub//atkinson/ENA.Materials/GUI/>, <http://www.math.uiowa.edu/~a>
- netlib Netlib: très nombreux programmes, <http://www.netlib.org/>
- fftw FFTW: transformée de Fourier rapide, <http://www.fftw.org/>
- NFFT From: Daniel Potts <potts@mathematik.tu-chemnitz.de>
Date: Tue, 21 Nov 2006 13:18:49 +0100
Subject: Software: NFFT, Nonequispaced Fast Fourier Transform

¹ACM= Association for Computing Machinery.

²SIAM = Society for Industrial and Applied Mathematics.

We are pleased to announce version 3.0 of our C library for computing the nonequispaced discrete Fourier transform in one or more dimensions. Our library is free software, and based on FFTW3. Visit the NFFT web-site at

<http://www.uni-chemnitz.de/~potts/nfft>

for the software, documentation, and related links.

- Laval Analyse numérique à l'univ. Laval³ <http://www.mat.ulaval.ca/anum/>
- NumRec W.H. Press, S.A. Teukolsky, W.A. Vetterling, B.P. Flannery, *Numerical Recipes*: programmes dans <http://nr.harvard.edu/numerical-recipes/>, texte dans <http://cfatab.harvard.edu/nr/nronline.html>, voir aussi <http://math.jpl.nasa.gov/nr/> pour une critique sévère et d'autres informations...
- GSL New Release of GNU Scientific Library. Version 1.6 of the GNU Scientific Library (GSL) is now available for download at:
<http://www.gnu.org/software/gsl/>
GSL is a free numerical library written in C using modern coding conventions. It is distributed under the GNU General Public License.
- CodeCogs Date: Tue, 15 Mar 2005 Subject: CodeCogs, Open Source C/C++ Library
This is to inform you about the website called "Code Cogs", which is "A New Open Source Scientific Library / Database in C++". The web site is free to use with all software being open source.
The main homepage is at: <http://www.codecogs.com>
CodeCogs is a new online scientific numerical repository of C/C++ code that has been created for scientists, engineers, mathematicians and financial workers.

³Merci à C. Debiève pour cette information.

Analyse numérique et théorie de l'approximation.

1. Qu'est ce que l'analyse numérique?

*Vaste programme.
Gal de Gaulle*

1.1. Analyse numérique et analyse.

Les mathématiciens du passé mêlaient allègrement les constructions calculatoires aux vastes généralisations. On pouvait passer du dessin de la coque d'un navire ou du volume des tonneaux de bière aux considérations les plus philosophiques. On est aujourd'hui plus disciplinaire, si pas plus discipliné, et on enferme la pensée dans toutes sortes de sous-régions.

L'*analyse* étudie les nombres réels et les fonctions de variables réelles. On considère, d'ailleurs à juste titre, que le progrès en analyse consiste à établir la vérité d'une proposition en utilisant un minimum de moyens. Ceci conduit à éliminer autant que possible toute construction détaillée, donc à éviter toute représentation explicite inutile. Ainsi, on préfère parler en analyse de formes ou d'opérateurs, voire d'endomorphismes, plutôt que de matrices qui renvoient à des représentations dans des bases imposées. A un niveau plus fondamental, utiliser la représentation décimale d'un nombre réel serait une lourde faute de goût.

L'*analyse numérique* est précisément liée à cette ancienne partie de l'analyse qui décrit ces constructions qui ne sont plus essentielles à la compréhension de la théorie. On ne se contente pas d'exhumer d'anciens secrets et de les rassembler dans de gros formulaires, ces constructions sont présentées de façon aussi ordonnée et systématique que possible: Henrici [**Hen**] définit l'analyse numérique comme la *théorie des méthodes constructives de l'analyse*, Natanson intitule son ouvrage [**Nat**] *Théorie constructive des fonctions*, Trefethen⁴ dit que l'analyse numérique est *l'étude des algorithmes de résolution des problèmes des mathématiques continues*.

L'analyse numérique commente, illustre, concrétise⁵ et applique l'analyse.

1.2. Analyse numérique et calcul.

Tout projet scientifique ou technique comporte une importante phase de calculs. Parlett⁶ décrit la tour

Programmes d'applications	5
Grands logiciels orientés vers les applications	4
Bibliothèques de programmes pour matrices, approximation, équadiff, optimisation, etc.	3
Fonctions élémentaires: $\log(x)$, etc.	2
Langages de programmation	1
Unité arithmétique + - × / Assembleur	0

du calcul scientifique (cf. figure de gauche) mise à la disposition de l'utilisateur. Chaque niveau utilise des niveaux inférieurs. L'analyse numérique intervient surtout aux niveaux 2 ("boîtes noires": problèmes banalisés) et 3, où les problèmes sont d'ailleurs explicitement formulés en termes mathématiques. On ne s'occupera donc pas beaucoup ici des représentations en machine des réels, ni de questions d'erreurs d'arrondi, qui ressortissent plutôt de l'algorithmique numérique. Quant aux niveaux supérieurs, ils intéressent spécifiquement des utilisateurs d'un domaine des sciences et techniques (niveau 4), et la mise au point d'une application particulière.

⁴ L.N. Trefethen, The Definition of Numerical Analysis, *SIAM News*, November 1992 = *Bulletin of the Institute of Mathematics and Applications*, March/April 1993 = <http://www.comlab.ox.ac.uk/oucl/users/nick.trefethen/defn.ps>, copie dans `trefethennuman.ps`

⁵L'analyse numérique ne remplace pas l'analyse par un projet où ne figureraient que des objets dûment construits (mathématique intuitionniste), toutes les méthodes de l'analyse sont parfaitement valides en analyse numérique, où on peut donc très bien trouver des démonstrations non constructives!

⁶B. Parlett, Progress in numerical analysis, *SIAM Review* **20** (1978) 443-456.

G. Strang distingue⁷ distingue six étapes dans le traitement d’un problème difficile:

- (1) **Modélisation.** Arriver au problème mathématique traduisant le mieux le phénomène considéré,
- (2) **Représentation.** Choisir une famille de fonctions susceptibles de bien approcher la solution de 1., et la **base** qui servira à cette représentation,
- (3) **Paramètres.** Bien choisir les degrés, points de grille ou d’interpolation, qui assureront une erreur suffisamment faible,
- (4) **Algorithme.** Décrire les étapes de calcul aboutissant à la solution numérique,
- (5) **Programmation.** Tenir compte des moyens de calcul disponibles,
- (6) **Visualisation.** La réponse peut se limiter à quelques nombres, mais peut nécessiter une présentation sous forme de tables, graphes, etc.

L’analyse numérique est concernée par les points (2), (3), et (4) pour une certaine part (l’algorithmique numérique pour une autre part).

2. Théorie de l’approximation.

Errare humanum est

2.1. Les trois niveaux d’une théorie de l’approximation. La théorie de l’approximation est la plus aimable des sciences exactes. Le droit à l’erreur y est autorisé, et même encouragé. Un objet f (nombre réel, fonction, opérateur, etc.) mathématiquement défini mais inaccessible à des représentations élémentaires, est approché par un objet plus simple p . Au premier niveau, il s’agit de construire une approximation acceptable p , par exemple $p = 0.3333$ si $f = 1/3$.

Jusque dans les années 1950, on voyait dans les formulaires la “série de Renard”, dite encore des “nombres normaux”, qui consiste en la suite des 10 puissances successives de 1.25 et qui donne très approximativement des constantes mathématiques, physiques et technologiques remarquables:

$2^{1/3}$	$\pi/2$	cm/in	π	in/dm	2π	g
$10^{1/10}$						π^2
1.25	1.6	2	2.5	3.15	4	5 6.3 8 10

Au deuxième niveau, on examine si une loi permet d’expliquer la forme de diverses approximations successives 0.33, 0.333, etc. A ce niveau, on ne s’occupe plus des approximations proprement dites. Au troisième niveau, on caractérise des classes d’objets f à partir de dispositifs initialement destinés à fournir des approximations.

Autre exemple moins trivial: soit $f = \sqrt{2}$, p est un nombre rationnel a/b avec $0 < b \leq n$, et la distance $|\sqrt{2} - a/b|$ la plus petite possible. On a donc défini une *meilleure approximation* dans un ensemble, pour une *distance* donnée. On fait varier n et on note les *nouvelles* fractions qui apparaissent. En fait, il suffit de prendre $m =$ entier le plus proche de $n\sqrt{2}$ et de conserver les fractions donnant une erreur plus petite que la plus petite erreur précédente⁸:

*C’est plein, plein de force,
et plein, plein de nombre là-dedans.*
Alfred Jarry⁹

$\sqrt{2} - 1/1$	=	0.4142135623730950488016887
$\sqrt{2} - 3/2$	=	-0.0857864376269049511983113
$\sqrt{2} - 4/3$	=	0.0808802290397617154683554
$\sqrt{2} - 7/5$	=	0.0142135623730950488016887
$\sqrt{2} - 17/12$	=	-0.0024531042935716178649779
$\sqrt{2} - 24/17$	=	0.0024488564907421076252181
$\sqrt{2} - 41/29$	=	0.0004204589248191867327232
$\sqrt{2} - 99/70$	=	-0.0000721519126192369125970
$\sqrt{2} - 140/99$	=	0.0000721482316809073875473

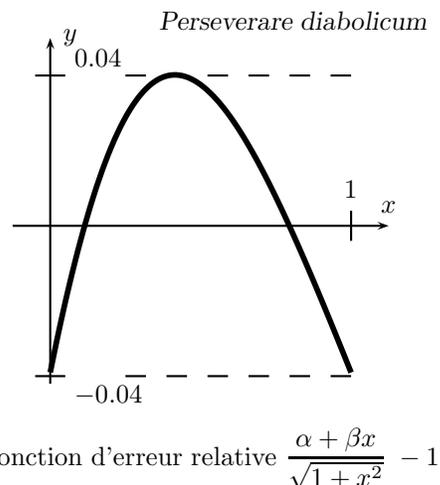
⁷ <http://www.ipam.ucla.edu/publications/inauguration/strang.html>

⁸On dispose donc déjà d’une approximation 1.4142135623730950488016887... de $\sqrt{2}$.

⁹dans *Le surmâle*, à propos d’un automate-boxeur de foire.

Au premier niveau, on recueille donc ces approximations. Au deuxième niveau, on examine comment décrire les approximations retenues¹⁰. Au troisième niveau, on apprend quelque chose sur $f = \sqrt{2}$, par exemple que 2 ne peut être le carré d'un nombre rationnel (voir p. 16), même si on connaît des démonstrations non constructives beaucoup plus courtes¹¹.

Autre exemple: Poncelet^a cherche à approcher $\sqrt{a^2 + b^2}$ par une expression plus simple $\alpha a + \beta b$. Il trouve $\alpha = 0.961$ et $\beta = 0.398$ assurant une erreur relative $\leq 4\%$ si on a choisi $a \geq b$: premier niveau. Mais comment a-t-il obtenu ces valeurs? Pour tout couple (α, β) , on considère l'erreur relative $\frac{\alpha a + \beta b}{\sqrt{a^2 + b^2}} - 1 = \frac{\alpha + \beta x}{\sqrt{1 + x^2}} - 1$, et on détermine pour quelles valeurs de $x = b/a \in [0, 1]$ l'erreur relative a la plus grande valeur absolue. On constate qu'il faut examiner trois valeurs de x , et que la plus grande erreur relative est minimisée quand les erreurs en ces trois valeurs de x ont une même valeur absolue et des signes alternés, principe qui sera considérablement développé par Tchebycheff (tout le chap. 2): deuxième niveau.



La meilleure fonction d'erreur relative $\frac{\alpha + \beta x}{\sqrt{1 + x^2}} - 1$

^aVoir V.L. Goncharov, The theory of best approximations of functions, *J. Approx. Theory* **106** (2000) 2-57.

Dernier exemple, relatif à l'analyse proprement dite cette fois, qui sera étudié dans ce cours (p. 60 et fin du chap. 3): des fonctions très lisses, disons \mathcal{C}^m sur un intervalle compact, sont susceptibles d'être approchées par des polynômes de degré $\leq n$ avec une erreur se comportant comme une puissance négative de n . Quelles sont les fonctions pour lesquelles l'erreur se comporte comme une puissance *faiblement* négative de n ? On obtient des fonctions continues partout et dérivables nulle part... , question d'analyse pure (troisième niveau) fort bien éclairée par ce recours à l'approximation.

”Je vis dans l'approximatif et je m'en rapproche de plus en plus”

Julos Beaucarne

3. Quelques approximations de fonctions utilisées dans les calculatrices et les ordinateurs

3.1. Calculatrices scientifiques: le système CORDIC.

Beaucoup de calculatrices scientifiques et les premiers coprocesseurs mathématiques ont comme opérations de base l'addition, la soustraction et le décalage (= multiplication ou division par 10). On accumule de telles opérations pour multiplier et diviser (comme “à la main”). On peut également incorporer la racine carrée à ce niveau.

Pour l'exponentielle et le logarithme, on dispose en mémoire les logarithmes $\lambda_k = \ln(1 + 10^{-k})$ pour suffisamment de valeurs de k . Calcul de $y = \ln x$: on se ramène à $1 < x < 10$, on part de $y = 1$, ensuite:

¹⁰Le secret consiste à ne retenir d'abord que le *dernier* élément des groupes de p consécutifs avec $f - p$ de même signe. Ce sous-ensemble d'approximations (*réduites*) $\{1/1, 3/2, 7/5, 17/12, 41/29, \dots\}$ présente une structure tout à fait intéressante (fraction continue: cf. C. Brezinski, *History of Continued Fractions and Padé Approximants*, Springer, 1991), I. Niven, *Irrational Numbers*, AMS & Wiley

¹¹Les premières preuves de transcendance de e et π étaient aussi basées sur des constructions d'approximations. Voir aussi p. 16.

```

pour k=0,1,2,... { x'=x*(1+10^(-k)) ;
                  tant que x'<10 {y=y-lambda_k;x=x';x'=x'*(1+10^(-k));
                  }
                }

```

Calcul de $y = e^x$: soit $x > 0$, on part de $y = 1$, et

```

pour k=0,1,2,... { x'=x-lambda_k;
                  tant que x'>0 {y=y*(1+10^(-k)); x=x'; x'=x'-lambda_k;
                  }
                }

```

Pour les fonctions trigonométriques, hyperboliques et leurs inverses, on utilise intelligemment les formules d'addition de ces fonctions. Il suffit de se munir d'une table supplémentaire constituée des $\alpha_k = \arctg(10^{-k})$. C'est le système CORDIC (*COordinate Rotation DIgital Computer*) réalisé par Volder en 1959, mais on fait remonter l'idée à Henry Briggs (1561-1631), l'un des inventeurs des logarithmes!

Cf: J. Laporte, Le secret des algorithmes, *L'Ordinateur Individuel* n°24, 1981, 89-92,

C.W. Schelin, Calculator function approximation, *Amer. Math. Monthly* **90** (1983) 317-325,
 cordic.txt,

<http://devil.ece.utexas.edu/cordic.html>,

Calcul informatique des fonctions trigonométriques, Julien ÉLIE, <http://www.trigofacile.com/maths/trigo/calcul/c>

3.2. Approximations polynomiales et rationnelles.

Les ordinateurs utilisent le plus souvent l'addition, la soustraction, la multiplication, la division, et souvent aussi la racine carrée de nombres "flottants" (mantisse fois base^{exposant}, presque toujours en binaire) comme opérations élémentaires. Le but est alors d'obtenir la précision demandée avec un minimum d'opérations, c'est à dire d'approcher la fonction par un polynôme ou une fonction rationnelle. Cet exercice n'est pas étranger à ce qu'il y a de plus classique dans l'analyse, un développement tronqué de Taylor servant souvent de point de départ.

Ainsi, Hastings [**Has**, pp.132–137 et 84] propose quelques approximations de $\arctg x$, $-1 \leq x \leq 1$: (cf. aussi [**Abr**, p.81])

$$\begin{aligned}
 0.995354 x - 0.288679 x^3 + 0.079331 x^5 & \text{ avec erreur } \leq 6.08 \cdot 10^{-4}, \\
 0.9992150 x - 0.3211819 x^3 + 0.1462766 x^5 - 0.0389929 x^7 & (8.14 \cdot 10^{-5}), \\
 0.9998660 x - 0.3302995 x^3 + 0.1801410 x^5 - 0.0851330 x^7 + 0.0208351 x^9 & (1.14 \cdot 10^{-5}), \\
 0.99997726x - 0.33262347x^3 + 0.19354346x^5 - 0.11643287x^7 + 0.05265332x^9 - 0.01172120x^{11} & (1.66 \cdot 10^{-6})
 \end{aligned}$$

qui ressemblent à des développements tronqués de la série de Taylor-Maclaurin $\arctg x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \dots$.

On verra comment construire de telles approximations. On devine qu'il faudra une dizaine de termes pour arriver à une erreur $\leq 10^{-8}$ (pour cette fonction, chaque fois que l'on ajoute un terme, l'erreur est à peu près divisée par $(\sqrt{2} + 1)^2$, on verra cela aussi).

D'autres formules plus économiques ont été imaginées, par exemple, d'après *VS FORTRAN Application Programming: Library Reference*, IBM Program Product SC26-3989, 1981¹², on se ramène d'abord à $0 \leq x \leq$

1, puis à $-\text{tg}(\pi/12) \leq x \leq \text{tg}(\pi/12) = 2 - \sqrt{3}$ par $\arctg x = \frac{\pi}{6} + \arctg \left(\frac{\sqrt{3}x - 1}{x + \sqrt{3}} \right)$ si $1 \geq x > \text{tg}(\pi/12)$,

enfin, l'approximation donnant $\arctg x$ dans $-\text{tg}(\pi/12) \leq x \leq \text{tg}(\pi/12)$ est

$$x \left[0.60310579 - 0.05160454x^2 + \frac{0.55913709}{x^2 + 1.4087812} \right]$$

avec une erreur $< 10^{-8}$.

Pour l'exponentielle, le même document propose

$$2^{-x} \approx 1 - \frac{2x}{0.034657359x^2 + x + 9.9545948 - \frac{617.97227}{x^2 + 87.417497}}$$

¹²Grand merci à M. J.L. Marrion qui a fourni cette documentation.

avec une erreur $< 10^{-8}$ quand on s'est ramené à $0 \leq x \leq 1$.

Des formules de même type sont utilisées pour \sin , \ln , etc.

Cf. aussi J.F. Hart & *al.*, *Computer Approximations*, Wiley, 1968.

3.3. AGM, etc.

Plus récemment, on a imaginé de nouvelles catégories d'algorithmes pour calculer des nombres remarquables en très haute précision (jusqu'à des *milliards* de décimales pour π).

L'algorithme **AGM** (Arithmetic-geometric mean) consiste, à partir de a_0 et b_0 , à calculer la suite de moyennes arithmétiques et géométriques successives $a_{n+1} = (a_n + b_n)/2$ et $b_{n+1} = \sqrt{a_n b_n}$. On calcule également les $c_{n+1} = (a_n - b_n)/2$ sous la forme $c_0 = \sqrt{a_0^2 - b_0^2}$, $c_{n+1} = c_n^2/(4a_{n+1})$. Gauss avait déjà montré que les suites $\{a_n\}$ et $\{b_n\}$ tendent très rapidement (quadratiquement) vers une limite commune $\text{AGM}(a_0, b_0)$

liée à une intégrale elliptique: $\frac{\pi}{2 \text{AGM}(a_0, b_0)} = \int_0^{\pi/2} \frac{d\theta}{\sqrt{a_0^2 - c_0^2 \sin^2 \theta}}$. L'usage de cet algorithme curieux est

longtemps resté confiné au calcul d'intégrales de ce type. Puis, R.P Brent (Fast multiple-precision evaluation of elementary functions, *J. ACM* **23** (1976), 242-251) et E. Salamin (Computation of π using arithmetic-geometric mean, *Math. Comp.* **30** (1976), 565-570) trouvèrent d'autres applications, à commencer par une formule rapide pour π :

$$\pi = \frac{2(\text{AGM}(1, 1/\sqrt{2}))^2}{1 - \sum_{i=0}^{\infty} 2^i c_i^2}.$$

De plus, $[c_n/(4a_n)]^{1/2^{n-1}}$ tend rapidement vers l'exponentielle de $-\pi \text{AGM}(a_0, b_0)/\text{AGM}(a_0, c_0)$, ce qui donne un moyen de calcul rapide des exponentielles et logarithmes, etc.

Cf. D.H. Bailey, Algorithm 719: Multiprecision translation and execution of FORTRAN programs, *ACM Trans. Math. Soft.* **19** (1993) 288-319; A FORTRAN 90- based multiprecision system, *ibid.* **21** (1995) 379-387.

J.M. Borwein, P.B. Borwein, The arithmetic-geometric mean and fast computation of elementary functions, *SIAM Rev.* **26** (1984) 351-366,

J.M. Borwein, P.B. Borwein, *Pi and the AGM*, Wiley, 1986,

Center for Experimental and Constructive Mathematics, Simon Fraser Univ., <http://www.cecm.sfu.ca/>

R. Preston, The mountains of pi, *The New Yorker*, 2 mars 1992.

3.4. Approximations et nombres irrationnels.

Les fractions $1/1, 3/2, 7/5, 17/12, 41/29, \dots$ apparaissant dans la note 10 de p. 14 sont y_n/x_n avec $y_n^2 - 2x_n^2 = (-1)^n$. On peut trouver une infinité de nombres entiers de plus en plus grands vérifiant cette égalité: en effet, si x_n, y_n conviennent, on constate que $x_{n+1} = x_n + y_n$ et $y_{n+1} = 2x_n + y_n$ donnent bien $y_{n+1}^2 - 2x_{n+1}^2 = -(y_n^2 - 2x_n^2)$. On voit¹³ aussi que cela revient à appliquer la méthode de la puissance $\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = A \begin{bmatrix} x_n \\ y_n \end{bmatrix}$ à la matrice $A = \begin{bmatrix} 1 & 1 \\ \sqrt{2} & 1 \end{bmatrix}$ et à obtenir des vecteurs de mieux en mieux alignés sur le vecteur propre $[1, \sqrt{2}]^T$: $y_{n+1} \pm \sqrt{2} x_{n+1} = (1 \pm \sqrt{2})(y_n \pm \sqrt{2} x_n)$. Donc, x_n et $y_n \rightarrow \infty$, $y_n - \sqrt{2} x_n \neq 0$ et $|y_n - \sqrt{2} x_n| \rightarrow 0$.

On en déduit l'irrationalité de $\sqrt{2}$ selon la proposition:

Si on trouve une infinité d'entiers x_n et y_n , avec $x_n \rightarrow \infty$, tels que $\alpha \neq y_n/x_n$ et $|\alpha x_n - y_n| \rightarrow 0$ quand $n \rightarrow \infty$, alors α est irrationnel.

(En effet, si $\alpha = A/B$ avec A et B entiers, on aurait $\alpha x_n - y_n = (Ax_n - By_n)/B = \text{entier}/B$, donc exactement zéro ou de valeur absolue bornée inférieurement par $1/|B|$. \square)

Ici, $|x_n \sqrt{2} - y_n| = |2x_n^2 - y_n^2|/(x_n \sqrt{2} + y_n) = 1/(x_n \sqrt{2} + y_n) \rightarrow 0$.

Des possibilités de montrer l'irrationalité de nombres remarquables sont donc liées à leurs facultés d'approximation.

Par exemple, $0 < \left| e - \frac{1}{1!} - \frac{1}{2!} - \dots - \frac{1}{n!} \right| < \frac{1}{(n+1)!} \frac{1}{1 - \frac{1}{n+2}} \Rightarrow e$ est irrationnel.

¹³Remarque aimablement communiquée par Michel Coyette.

On appelle *constante de Markov* (ou de *Lagrange*) d'un réel α l'expression $M(\alpha) = 1 / \left[\liminf_{\substack{x,y \in \mathbb{Z} \\ x \rightarrow \infty}} x^2 |\alpha - y/x| \right]$,

applications en equadiff et mécanique céleste¹⁴.

Les réduites successives $\frac{Y_1}{X_1} = \frac{a_1}{b_1}, \frac{Y_2}{X_2} = \frac{a_1}{b_1 + \frac{a_2}{b_2}}, \frac{Y_3}{X_3} = \frac{a_1}{b_1 + \frac{a_2}{b_2 + \frac{a_3}{b_3}}}$ etc. de la *fraction continue* $\alpha = \frac{a_1}{b_1 + \frac{a_2}{b_2 + \dots}}$

vérifient $X_{n+1} = b_{n+1}X_n + a_{n+1}X_{n-1}, Y_{n+1} = b_{n+1}Y_n + a_{n+1}Y_{n-1}$, et

$\frac{Y_{n+1}}{X_{n+1}} - \frac{Y_n}{X_n} = \frac{(-1)^n a_1 a_2 \dots a_{n+1}}{X_n X_{n+1}}$. Si on trouve Z_n tel que $X_n Z_n$ et $Y_n Z_n$ sont entiers, et si la fraction continue converge, on peut parfois tirer une condition d'irrationalité de

$$|\alpha Z_n X_n - Z_n Y_n| \leq Z_n X_n \left| \sum_{k=n}^{\infty} \frac{(-1)^k a_1 a_2 \dots a_{k+1}}{X_k X_{k+1}} \right|$$

Quelques fractions continues remarquables de *fonctions*:

$$\operatorname{tg} x = \frac{x}{1 - \frac{x^2}{3 - \frac{x^2}{5 - \frac{x^2}{7 - \dots}}}}, \quad e^x = 1 + \frac{x}{1 - \frac{x}{2 + \frac{x}{3 - \frac{x}{2 + \frac{x}{5 - \frac{x}{2 + \frac{x}{7 - \frac{x}{2 + \dots}}}}}}}}$$

permettent d'établir que $\operatorname{tg} x$ et e^x sont irrationnels quand x est rationnel $\neq 0$. On en déduit que π est irrationnel: sinon, $\operatorname{tg} \pi$ serait irrationnel...

Mais comment obtient-on de telles formules? Lagrange, Euler¹⁵: si $y = f(x)$ vérifie l'équation de *Riccati* $xa(x)y' = b(x)y^2 + c(x)y + d(x)$, où $a, b, c \in \mathcal{P}_m$, et $d \in \mathcal{P}_{m-1}$, alors, si $f(x) = \frac{f(0)}{1 - xf_1(x)}$, f_1 vérifie une équation de même type, avec les mêmes degrés. En effet, on a $xa(x)[xf_1' + f_1] = b(x)f(0) + c(x)[1 - xf_1(x)] + d(x)[1 - xf_1(x)]^2/f(0)$, d'où une équation de Riccati pour f_1 avec le même $a(x)$, $b_1(x) = xd(x)/f(0)$, $c_1(x) = -c(x) - 2d(x)/f(0) - a(x)$, et $d_1(x) = [b(x)f(0) + c(x) + d(x)/f(0)]/x$ (qui est bien un polynôme si f est développable en Taylor-Maclaurin). On itère ensuite $f_n(x) = \frac{f_n(0)}{1 - xf_{n+1}(x)}$, $n = 0, 1, \dots$: $b_{n+1}(x) = xd_n(x)/f_n(0), c_{n+1}(x) = -c_n(x) - 2d_n(x)/f_n(0) - a(x), d_{n+1}(x) = [b_n(x)f_n(0) + c_n(x) + d_n(x)/f_n(0)]/x, f_{n+1}(0) = -d_{n+1}(0)/c_{n+1}(0)$. Evidemment, ça marche si $f_n(0) \neq 0, n = 0, 1, \dots$

Pour la tangente, partir de $f_0(x) = \frac{\operatorname{tg} \sqrt{x}}{\sqrt{x}}$, alors, $a(x) = 2, b_0(x) = x, c_0 = -1, d_0 = 1$. On trouve pour $n > 0$, $b_n(x) = (2n - 1)x, c_n = -(2n + 1), d_n = 1/(2n - 1), f_n(0) = 1/(4n^2 - 1)$.

¹⁴K. Ben-Naoum & J. Mawhin, The periodic Dirichlet problem for some semilinear wave equations, *J. Diff. Eq.* **96** (1992) 340–354; A.K. Ben-Naoum, On the Dirichlet problem for the nonlinear wave equation in bounded domains with corner points, *Bull. Belg. Math. Soc. Simon Stevin* **3** (1996) 345–361; A.K. Ben-Naoum, Some results on the Markov number from the theory of Diophantine approximations, *Rapports Séminaire Math. UCL* n°254 (1996). voir aussi *Orthogonal Polynomials and Continued Fractions: From Eulers Point of View*, par Sergey Khrushchev, Cambridge University Press

¹⁵A.N. Khovanskii, *The Application of Continued Fractions and their Generalization to Problems in Approximation Theory*, P. Noordhoff, Groningen, 1963., D.E. Knuth, *The Art of Computer Programming*, vol. **2**, *Seminumerical Algorithms*, 3rd ed., Addison Wesley 1998, exercice 16, p. 375 (360 dans l'édition de 1981), et pp. 650-651 [602-603]

Enfin, $(2n - 1)f_n(x) = \frac{1}{2n + 1 - x(2n + 1)f_{n+1}(x)}$.

Pour l'exponentielle, on transforme à partir de $\frac{e^x - 1}{e^x + 1} = \tanh(x/2)$.

L'établissement de critères d'irrationalité et de transcendance à partir d'approximations de fonctions est une manifestation remarquable du troisième niveau de la théorie de l'approximation.

Dernier résultats sensationnels trouvés à ce jour: en 1979, R. Apéry montrait que $\zeta(3) = \sum_1^\infty n^{-3}$ est irrationnel au moyen de $n^3X_n - (34n^3 - 51n^2 + 27n - 5)X_{n-1} + (n-1)^3X_{n-2} = 0$ et $n^3Y_n - (34n^3 - 51n^2 + 27n - 5)Y_{n-1} + (n-1)^3Y_{n-2} = 0$, $X_0 = 1, X_1 = 5, Y_0 = 0, Y_1 = 6$. Une démonstration fait appel aux polynômes de Legendre, cf. M. Prevost, A new proof of the irrationality of $\zeta(2)$ and $\zeta(3)$ using Padé approximants, *J. Comp. Appl. Math.* **67** (1996) 219-235.

T. Rivoal, La fonction zêta de Riemann prend une infinité de valeurs irrationnelles aux entiers impairs. *C. R. Acad. Sci. Paris* **331** (2000), 267-270. T. Rivoal, Au moins un des neuf nombres $\zeta(5), \zeta(7), \dots, \zeta(21)$ est irrationnel. *Acta Arith.* **103** (2002), 157-167. <http://www-fourier.ujf-grenoble.fr/~rivoal/>

Voir aussi de nombreuses publications dans <http://www.cecm.sfu.ca/>

3.5. Approximations les plus simples: bien commencer. Martin Rebas, auteur du document "Approximations of Natural Numbers",

<http://www.rebas.se/humor/natural.shtml>, (copié dans

[approxnat.html](#)), qui est aussi le président du club des gens-qui-ne-connaissent-que-deux-décimales-de-pi, a jugé bon d'aider le débutant en fournissant des approximations des objets mathématiques les plus simples qui soient, les nombres entiers:

$$2 \approx \pi - \sqrt{e - \sqrt{2}}, 3 \approx \sqrt{e + \pi + \pi}, 4 \approx \frac{\sqrt{2}}{\sqrt{e}} + \pi, 5 \approx \sqrt{2}\sqrt{\pi - \sqrt{2}} + \pi, 6 \approx \frac{1}{\sqrt{\pi}} + e + e,$$

$$7 \approx \sqrt{\sqrt{1+e} - \sqrt{2}} + \pi + \pi, 8 \approx e\sqrt{2}\sqrt{\sqrt{2} + \pi}, 9 \approx e + \pi + \pi, 10 \approx \sqrt{2} \left(\frac{\pi-1}{e} + \pi + \pi \right).$$

CHAPITRE 1

Théorèmes généraux d'existence et d'unicité de meilleure approximation.

Pour un élément f d'un ensemble X , il est question d'examiner si on peut trouver $p \in V \subset X$ le plus proche de f .

1. Distances et normes.

1.1. Rappelons qu'une *distance* sur un ensemble X est une fonction d définie pour tout couple d'éléments de X , avec

- (1) $\forall (f, g) \in X \times X, d(f, g) \geq 0$,
- (2) $(f = g) \iff (d(f, g) = 0)$,
- (3) $\forall (f, g) \in X \times X, d(f, g) = d(g, f)$ (symétrie),
- (4) $\forall (f, g, h) \in X \times X \times X, d(f, g) \leq d(f, h) + d(h, g)$ (inégalité triangulaire).

On appelle *espace métrique* tout ensemble X muni d'une distance.

1.2. Exercices et exemples. Exemples triviaux, grands cercles sur la sphère, distance cordale sur la sphère, sur $\overline{\mathbb{C}}$ identifié à la sphère de Riemann, géodésiques sur une variété, demi-plan de Poincaré, lire les pages 65-68 de H. Poincaré: *La science et l'hypothèse*¹, Flammarion, 1902 = "Champs Flammarion", 1968. Distance p -adique sur \mathbb{Q} .

$$\text{Distance de Hausdorff } D(A, B) = \max \left\{ \sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(a, b) \right\}.$$

Distance de Mahalanabis dans \mathbb{R}^n : $d(\mathbf{x}, \mathbf{y}) = [(\mathbf{x} - \mathbf{y})^T \mathbf{V}^{-1}(\mathbf{x} - \mathbf{y})]^{1/2}$, où \mathbf{V} est une matrice de variance-covariance.

1.3. Remarques. Choquet (*Cours d'analyse II, topologie*, Masson, 1969) définit également (pp.60-62) les **écarts** et les **jauges**.

On est très habitué maintenant à voir ces notions topologiques exprimées en termes d'inégalités non strictes \leq et \geq . Il n'en a pas toujours été ainsi: Laurent Schwartz est conscient d'innover en écrivant² dans son *Cours d'analyse* (Hermann, 1967) à la page 17:

"Notons que nous rompons ici avec l'usage antérieurement acquis en appelant inférieur ce qu'on appelait inférieur ou égal, et strictement inférieur ce qu'on appelait inférieur³. La raison d'être de ces changements, pleinement justifiés par la suite, est que la notion la plus généralement utilisée est \leq , et qu'il est bon qu'elle ait l'appellation la plus courte. On devra toujours utiliser \leq plutôt que $<$, toutes les fois que cela sera possible; quand on écrira une inégalité stricte avec $<$, ce sera pour avertir le lecteur qu'il y a un point délicat, et que l'inégalité large \leq ne conviendrait pas. Par exemple, la continuité d'une fonction réelle d'une variable réelle en un point a s'écrira ainsi:

$$\text{quel que soit } \varepsilon > 0, \text{ il existe } \eta > 0 \text{ tel que } |x - a| \leq \eta \text{ entraîne } |f(x) - f(a)| \leq \varepsilon$$

¹Voir <http://cedric.cnam.fr/ABU/BIB/auteurs/poincareh.html>

²Remarque aimablement communiquée par J. Meinguet.

³En anglais, \leq et \geq sont toujours "less or equal" et "greater or equal" (note de l'A.)

Nous avons mis le symbole d'inégalité large toutes les fois que c'était possible, et nous n'avons employé l'inégalité stricte > 0 que là où c'était absolument nécessaire à l'énoncé."

*Sans transition, la suite
PPDA*

1.4. Normes. Une *norme* sur un espace *vectoriel* X (sur \mathbb{R} ou \mathbb{C}) doit vérifier

- (1) $\forall f \in X, \|f\| \geq 0,$
- (2) $f = 0 \iff \|f\| = 0,$
- (3) $\|\alpha f\| = |\alpha| \|f\|,$
- (4) $\forall f, g \in X, \|f + g\| \leq \|f\| + \|g\|.$

Alors, $\|f - g\|$ est une distance valide.

1.5. Exemples, exercices.

Montrez que $\|f \pm g\| \geq \|f\| - \|g\|$ (dans tout triangle, tout côté est supérieur à la *différence* des deux autres).

Normes de Hölder sur \mathbb{R}^m ou \mathbb{C}^m :

$$\|x\|_p = (\sum_1^m |x_k|^p)^{1/p}, 1 \leq p \leq \infty \quad (\|x\|_\infty = \max_k |x_k|). \quad \text{Cf. [Dav] pp. 130–133.}$$

Normes de Hölder sur $L^p(A)$:

$$\|f\|_p = (\int_A |f(t)|^p)^{1/p}, 1 \leq p \leq \infty \quad (\|f\|_\infty = \sup_{t \in A} |f(t)|).$$

Normes pondérées: si $w_k > 0, k = 1, \dots, m,$ $\|x\|_w = \|y\|,$ avec $y_k = w_k x_k$ est encore une norme; si $w(t) > 0$ presque partout sur $A,$ $\|w(t)f(t)\|$ est encore une norme de f .

La norme est une fonction continue sur X : si $\lim_{n \rightarrow \infty} f_n = f,$ c'est-à-dire si $\|f - f_n\| \rightarrow 0$ quand $n \rightarrow \infty,$ on a $\|f_n\| \rightarrow \|f\|,$ puisque $|\|f\| - \|f_n\|| = |\|f_n + (f - f_n)\| - \|f_n\|| \leq \|f - f_n\|.$

La norme est même une fonction lipschitzienne: $\left| \frac{\|f\| - \|g\|}{\|f - g\|} \right| \leq L = 1.$

Sur un espace de *dimension finie* $V,$ la norme est une fonction continue des coordonnées dans une base quelconque: soient $\{a_1, \dots, a_n\}$ et $\{b_1, \dots, b_m\}$ les composantes de f et g dans V de base $\varphi_1, \dots, \varphi_m\}.$ Alors,

$$\|f - g\| = \left\| \sum_1^k (a_k - b_k) \varphi_k \right\| \leq \sum_1^k |a_k - b_k| \|\varphi_k\| \leq C \max_k |a_k - b_k|,$$

avec $C = \sum_1^k \|\varphi_k\|.$

Inversement, les composantes sont continues selon toute norme donnée sur V de dimension finie: si $f = \sum_1^m a_j \varphi_j$ et $g = \sum_1^m b_j \varphi_j,$ montrons que $|a_k - b_k| \leq C' \|f - g\|,$ où C' ne dépend que de la base $\{\varphi_1, \dots, \varphi_m\}$ de $V.$ Sinon, on pourrait trouver une infinité d'éléments $f_n \neq 0$ de V avec $f_n = \sum_1^m a_j^{(n)} \varphi_j,$ et $|a_k^{(n)}| / \|f_n\| \rightarrow \infty.$ Divisons les composantes de chaque f_n par la plus grande d'entre elles: on a maintenant une suite $\{g_n\}$ de composantes bornées (par l'unité), avec $\|g_n\| \rightarrow 0.$ Extrayons des suites convergentes de composantes, on obtient une suite $\{g_{n_i}\}_i$ convergente, et on peut s'arranger pour que l'une des composantes de la limite soit égale à 1. Comme $\|g_{n_i}\| \rightarrow 0,$ on a une représentation de $\lim_{n \rightarrow \infty} g_{n_i} =$ le vecteur nul avec des composantes non nulles, ce qui est impossible si $\{\varphi_k\}$ est une base de $V.$

Tout ceci confirme la proposition bien connue *d'équivalence des normes sur un espace vectoriel de dimension finie*: si $\|\cdot\|_\alpha$ et $\|\cdot\|_\beta$ sont deux normes valides sur V de dimension finie,

$$\forall f \in V, D \|f\|_\beta \leq \|f\|_\alpha \leq C \|f\|_\beta,$$

avec $0 < D \leq C < \infty.$

Si $\|\cdot\|_\alpha$ est une norme donnée, et si $\|\cdot\|_\beta$ est une norme, utile aux calculs, construite à partir d'une représentation dans une base, la base est d'autant plus intéressante que le rapport C/D

est petit (proche de l'unité). En effet, pour estimer la proximité de deux éléments de X en erreur relative $\|f - g\|_\alpha / \|f\|_\alpha$ en ne connaissant que la valeur des normes $\|\cdot\|_\beta$, on a

$$\frac{1}{K} = \frac{D}{C} \leq \frac{\|f - g\|_\alpha / \|f\|_\alpha}{\|f - g\|_\beta / \|f\|_\beta} \leq \frac{C}{D} = K \quad (\text{nombre de condition, ou conditionnement}) \quad (1)$$

Savoir que $C/D < \infty$, c'est de l'analyse; savoir estimer C/D , c'est de l'analyse numérique.

Notation. On désigne par \mathcal{P}_n l'ensemble des polynômes de degré $\leq n$. C'est une espace vectoriel de dimension $n + 1$.

N.B.: L'ensemble des polynômes de degré *exact* n n'est pas un espace vectoriel!

1.6. Exercices, exemples. On prend $\|\cdot\|_\alpha$ = norme du maximum pour des fonctions continues sur $[0, 1]$, $V = \mathcal{P}_1, \mathcal{P}_2, \dots$, $\|\cdot\|_\beta$ = maximum des |coefficients| dans la base des monômes $1, x, x^2, \dots$. Ce n'est justement pas une très bonne base quand le degré augmente:

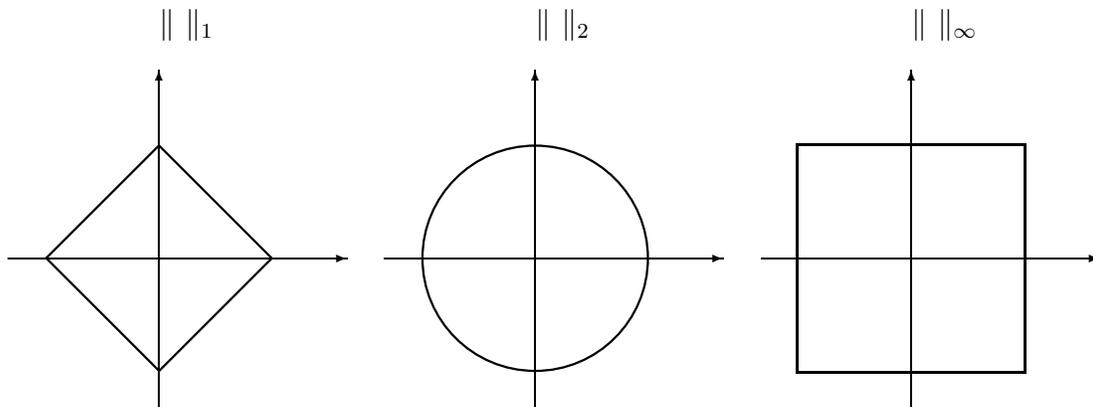
$$\frac{1}{2} \max(|a|, |b|) \leq \max_{0 \leq x \leq 1} |ax + b| \leq 2 \max(|a|, |b|),$$

$$\frac{1}{8} \max(|a|, |b|, |c|) \leq \max_{0 \leq x \leq 1} |ax^2 + bx + c| \leq 3 \max(|a|, |b|, |c|),$$

comment les choses évoluent-elles avec le degré? (voir plus loin (point 3, p. 42).

On notera B_r la boule $\{f \in X : \|f\| \leq r\}$ (fermée) de centre 0 et de rayon r , $B_r(c)$ la boule $\{f \in X : \|f - c\| \leq r\}$ de centre $c \in X$ et de rayon r .

Formes de la boule unité B_1 de \mathbb{R}^2 :



1.7. Formes et applications linéaires continues sur des espaces vectoriels normés de fonctions.

1.7.1. L'expression générique d'une forme continue sur l'espace $\mathcal{C}[a, b]$ des fonctions continues sur $[a, b]$, normé par $\|\cdot\|_\infty$, est

$$\varphi : \varphi(f) = \int_a^b f(x) \psi(x) dx,$$

où ψ est une fonction intégrable sur $[a, b]$. On établit d'ailleurs que la norme de cette forme φ vaut

$$\|\varphi\| = \sup_{\|f\|_\infty \leq 1} \left| \int_a^b f(x) \psi(x) dx \right| = \int_a^b |\psi(x)| dx = \|\psi\|_1.$$

Ce ne sont pas les seules expressions possibles: toute combinaison de valeurs ponctuelles

$$\varphi : \varphi(f) = \sum_{k=1}^M \alpha_k f(x_k), \quad (2)$$

où les α_k et les x_k sont fixés, convient également. On trouve alors (c'est encore plus facile),

$$\|\varphi\| = \sum_{k=1}^M |\alpha_k|.$$

Voir aussi les intégrales de Riemann-Stieltjes $\int_a^b f(x) d\mu(x)$ au chap. 3, (53), p. 75.

1.7.2. . Une forme contenant des dérivées, comme

$$\varphi(f) = \int_a^b f'(x) \psi(x) dx, \quad \text{ou } \varphi(f) = \sum_{k=1}^M \alpha_k f'(x_k),$$

n'est normalement pas bornée dans un espace de fonctions normé par $\|\cdot\|_\infty$. On doit prendre une autre norme, par exemple $\|f\| = \|f\|_\infty + \|f'\|_\infty$.

1.7.3. . Dans un espace X de dimension *finie* de fonctions continues, toute forme définie partout est continue et peut d'ailleurs s'exprimer comme (2):

si $\{b_0, \dots, b_N\}$ est une base de X , on peut écrire, $\forall f \in X$, $f = \sum_0^N c_i b_i$. Alors, on voit que φ ne dépend que des $\varphi(b_i)$: $\varphi(f) = \sum_0^N c_i \varphi(b_i)$.

Éliminons les c_i à partir de $N+1$ valeurs de f en des points x_0, \dots, x_N : $f(x_i) = \sum_0^N c_j b_j(x_i)$, d'où

$$[\alpha_0, \dots, \alpha_N] = [\varphi(b_0), \dots, \varphi(b_N)] [b_j(x_i)]^{-1},$$

valable si la matrice des $b_j(x_i)$ est non singulière (**unisolvance**, voir plus loin, problème d'interpolation en général, § 1.2, p. 146).

1.7.4. . Dans L^p , utiliser Hölder:

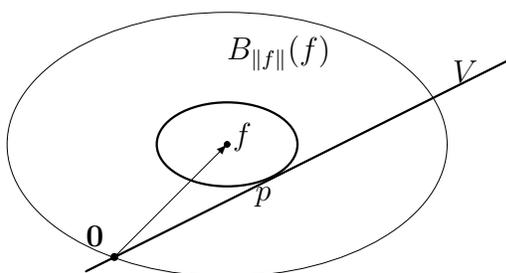
$$\varphi(f) = \int_a^b f(x) \psi(x) dx \Rightarrow \|\varphi\| = \|\psi\|_q,$$

si $\psi \in L^q$, $p^{-1} + q^{-1} = 1$.

2. Existence d'une meilleure approximation.

Pouvez-vous dire mieux?
Coluche

2.1. Théorème d'existence de meilleure approximation dans un sous-espace de dimension finie. Soit X un espace vectoriel normé sur \mathbb{R} ou \mathbb{C} et V un sous-espace vectoriel de dimension finie de X . Alors, $\forall f \in X$, il existe au moins un $p \in V$ tel que $\|f - p\| = \inf_{q \in V} \|f - q\|$.



Démonstration: en effet, soit $E = \inf_{q \in V} \|f - q\|$. Comme $q = 0$ est dans V , on a $E \leq \|f\|$. Il suffit donc d'examiner $\|f - q\|$ sur la partie *fermée bornée* $K = B_{\|f\|}(f) \cap V$ de V de dimension finie. La fonction continue $\|f - q\|$ atteint son infimum E en au moins un point $q = p$ du compact K . \square

Sur la figure, on a noté $B_{\|f\|}(f)$ qui contient 0 sur sa frontière, et aussi $B_E(f)$ (en trait plus gras) qui touche V en p .

La démonstration peut se faire en termes des composantes de q dans une base de V . La norme $\|f - q\|$ apparaît alors comme une fonction continue de ces composantes réelles ou complexes, que l'on minimise donc sur une partie compacte \tilde{K} de \mathbb{R}^m ou \mathbb{C}^m , avec $\tilde{K} = \{a_1, \dots, a_m : \sum_1^m a_k \varphi_k \in K\}$.

Remarquons que K est entièrement contenu dans la boule $B_{2\|f\|}$: si $\|f - q\| = \|q - f\| \leq \|f\|$, $\|q\| = \|f + (q - f)\| \leq \|f\| + \|q - f\| \leq 2\|f\|$. Le vecteur des composantes de q vérifie donc $\|a\| \leq 2\|f\|/D$.

2.2. Contre-exemple. Voici un exemple de non-existence quand on approche sur une partie de X qui n'est pas un sous-espace vectoriel: on prend $X = \mathcal{C}[-1, 1]$, la norme du maximum, $f(t) = t$, $V = \{ae^{bt} + ce^{dt}\}_{a,b,c,d \in \mathbb{R}}$. On trouve $E = 0$: prendre $q(t) = N(e^{t/N} - e^{-t/N})/2 = N \sinh(t/N) = t + t^3/(6N^2) + \dots$, on a donc $\|t - q(t)\|_\infty$ aussi petit que l'on veut, mais il n'existe pas de $q \in V$ tel que $\|q - f\| = E = 0$, $f(t) = t$ n'est pas une combinaison d'exponentielles.

Cf. aussi [Dav, p.153].

Autre exemple de supremum inaccessible: soit F une fonction positive strictement décroissante sur $[0, \infty)$. Problème 1: maximiser la moyenne de F , $\int_0^\infty F(x) d\mu(x)$ sur les mesures de probabilité μ . Le résultat ne peut dépasser $F(0)$ et est évidemment réalisé avec la mesure ponctuelle δ_0 en 0. Problème 2: on se limite maintenant aux mesures de moyenne $m > 0$. On peut encore s'approcher autant que l'on veut de $F(0)$: prendre les mesures $(1-\varepsilon)\delta_0 + \varepsilon\delta_{m/\varepsilon}$ qui ont bien une moyenne $\int x d\mu(x)$ égale à m , et qui donnent $\int F d\mu = (1-\varepsilon)F(0) + \varepsilon F(m/\varepsilon)$. Le supremum est donc encore $F(0)$. Mais il n'est atteint que par δ_0 qui a une moyenne $0 \neq m \dots$

2.3. Remarque. Si p est une meilleure approximation de f dans V et si $q \in V$, $p + q$ est une meilleure approximation de $f + q$ dans V . Ou encore: $f - p$ est l'élément de **norme minimale** de **l'espace affine** $f + V$.

En effet, soit $E = \|f - p\|$. On a $\|f + q - (p + q)\| = \|f - p\| = E$. S'il y avait un élément r de V vérifiant $\|f + q - r\| < E$, $r - q$ serait une meilleure approximation de f que p .

3. Unicité de la meilleure approximation.

3.1. Définition. Convexité. Une partie P d'un espace vectoriel X est *convexe* si f et $g \in P \implies \lambda f + (1 - \lambda)g \in P$ pour $\forall \lambda \in [0, 1]$ (*combinaison convexe* de f et g). Toute boule $B_r(f) = \{g \in X : \|g - f\| \leq r\}$ d'un espace métrique X est convexe. Toute variété linéaire de X est convexe.

3.2. Proposition. Dans les conditions du théorème précédent, l'ensemble des meilleures approximations de f dans V est une partie convexe de V .

En effet, cet ensemble est $B_E(f) \cap V$, l'intersection de deux convexes est convexe. Par exemple, prendre la norme du maximum sur $[-1, 1]$, $f(t) = t$, $V = \{at^2 + b\}_{a,b \in \mathbb{R}}$.

3.3. Définition. Une norme est *stricte* si la boule unité est strictement convexe, c'est-à-dire si $\|f\| \leq 1, \|g\| \leq 1, \|f + g\| = 2 \Rightarrow f = g$. Les normes $\|\cdot\|_1$ et $\|\cdot\|_\infty$ ne sont pas strictes. La norme $\|\cdot\|_p$ est stricte si $1 < p < \infty$ ([Dav, p.141]).

On verra plus tard (chap. 3) qu'une norme dérivant d'un produit scalaire est toujours stricte: $\|f + g\|^2 = \|f\|^2 + 2(f, g) + \|g\|^2$ ne vaut 4 que si $\|f\| = \|g\| = (f, g) = 1 \Rightarrow \|f - g\|^2 = \|f\|^2 - 2(f, g) + \|g\|^2 = 0$.

3.4. Une condition suffisante d'unicité. Théorème. *Le problème posé au théorème 2.1 a une solution unique si la norme de X est stricte.*

En effet, toujours avec $E := \min_{q \in V} \|f - q\|$, soit $E > 0$ (si $E = 0$, la seule meilleure approximation est f lui-même). Supposons que p_1 et p_2 sont deux meilleures approximations distinctes de f . On a $\left\| \frac{f - p_1}{E} \right\| = 1, \left\| \frac{f - p_2}{E} \right\| = 1$, donc $\left\| \frac{2f - p_1 - p_2}{E} \right\| < 2$, ou encore $\|f - (p_1 + p_2)/2\| < E$: $(p_1 + p_2)/2$ serait un élément de V meilleur que p_1 ou p_2 . \square

Dans le cas des normes non strictes, des théorèmes d'unicité pourront encore être établis, on verra un cas au chapitre suivant.

3.5. Exercice. En plus de l'exemple dans 3.2, voir ces deux exemples de non unicité ([Pow, pp. 18-19]): $X = \mathcal{C}[-1, 1]$ avec $\|\cdot\|_1, f =$ la constante 1, $V = \{\lambda x\}_{\lambda \in \mathbb{R}}$; le même X , mais avec $\|\cdot\|_\infty$, le même $f, V = \{\lambda(1 + x)\}_{\lambda \in \mathbb{R}}$.

4. Continuité du projecteur de meilleure approximation.

Que la norme soit stricte ou non, si X et V sont tels que $\forall f \in X$ possède exactement une meilleure approximation dans V , on peut définir le projecteur P :

$$Pf = \text{meilleure approximation de } f \text{ dans } V.$$

Ce projecteur ($P^2 = P$) est normalement une application *non linéaire* $X \rightarrow V$.

Par exemple, prendre $X = \mathcal{C}[-1, 1]$, la norme du maximum, $V =$ constantes, $f(t) = t, g(t) = t^2$, on a $Pf = 0, Pg = 1/2, P(f + g) = 7/8 \neq Pf + Pg$. (Sur les constantes, $Pf = (\max_{t \in A} f(t) + \min_{t \in A} f(t))/2$.)

On a quand même $P(\alpha f) = \alpha Pf$ pour tout scalaire α .

4.1. Théorème de continuité. *Dans les conditions du théorème 2.1, si tout f de X admet exactement une meilleure approximation Pf dans V , l'application P est continue sur X .*

En effet ([Che] pp. 23–24), soit $f \in X$. Il faut montrer que, $\forall \epsilon > 0$, il existe un voisinage non vide $\{g : \|g - f\| \leq \delta\}$ entièrement envoyé par P dans la boule de centre Pf et de rayon ϵ .

Supposons qu'il n'en est rien: il existe donc un $\epsilon > 0$ tel que tous les voisinages de f ont des images par P contenant au moins un élément situé à distance $> \epsilon$ de Pf ...

Il y a donc une suite $\{f_n\}$ convergeant vers f : $f_n \rightarrow_{n \rightarrow \infty} f$ dans X , et dont les images ne tendent pas vers Pf : $\|Pf_n - Pf\| > \epsilon$ fixé > 0 .

Comme $\|f_n\| \rightarrow \|f\|$ (continuité de la norme), toutes les normes $\|f_n\|$ sont bornées par un même nombre M ; on a vu que Pf_n se trouve nécessairement dans la boule de centre f_n et de rayon $\|f_n\|$, donc dans $B_{2\|f_n\|} \subseteq B_{2M}$. Tous les Pf_n et Pf sont donc dans le compact $B_{2M} \cap V$ de V (compact parce que V est de dimension finie). Nous pouvons donc considérer une suite extraite **convergente** $\{Pf_{n_i}\}$ des Pf_n , de limite $h \in V$, nécessairement distinct de Pf , mais

alors

$$\begin{aligned} \|h - f\| &= \lim_{j \rightarrow \infty} \|Pf_{n_j} - f_{n_j}\| \\ &\leq \lim_{j \rightarrow \infty} \|Pf - f_{n_j}\| \quad \text{puisque } Pf_{n_j} \text{ est meilleure que } Pf \\ &\leq \|Pf - f\| \end{aligned}$$

impossible si $h \neq Pf$, puisque Pf est la *seule* meilleure approximation de f dans V . \square

4.2. Forte unicité. On peut parfois quantifier la distance entre f et un élément quelconque q de V par une inégalité de la forme

$$\|f - q\| \geq \|f - Pf\| + \gamma \|Pf - q\|,$$

avec $\gamma > 0$ dépendant de f et V , mais pas de q . Cela revient à une condition de Lipschitz pour le projecteur P :

$$\frac{\|Pf - Pg\|}{\|f - g\|} \leq \lambda,$$

avec $\lambda = 2/\gamma$ ([Che] pp. 80–82).

Surveys in Approximation Theory (SAT), Volume 5 (2010)

* 12. András Kro and Allan Pinkus, Strong Uniqueness, 1-91. <http://www.emis.de/journals/SAT/papers/12>

5. Dualité.

Si λ est une forme linéaire $\neq 0$ continue sur X , de *noyau contenant* V , on a $\lambda(f) = \lambda(f - q)$ pour $\forall q \in V$, en particulier avec la meilleure approximation p de f :

$$|\lambda(f)| = |\lambda(f - p)| \leq \|\lambda\| E(f),$$

où $E(f) := \|f - p\|$. On a donc la *borne inférieure* $E(f) \geq |\lambda(f)|/\|\lambda\|$ si $\lambda(q) = 0, \forall q \in V$.

On aura l'occasion de concrétiser cette propriété au chapitre suivant (§§ 2.2 et 2.4).

Théorème. *Si X est un espace de Banach, V un sous-espace de dimension finie de X ,*

$$E(f) = \max_{\substack{\lambda \in V^\perp \\ \|\lambda\|=1}} \lambda(f).$$

On applique le théorème de Hahn-Banach ([DeVLor] p.61, [Mei] § 1.3).

Extension à la meilleure approximation sur une partie convexe de X :

Théorème de Fenchel. *Si W est une partie convexe de l'espace normé X , on a*

$$E(f) := \inf_{q \in W} \|f - q\| = \sup_{\substack{\lambda \in X^* \\ \|\lambda\| \leq 1}} \left[\lambda(f) - \sup_{q \in W} \lambda(q) \right].$$

Cf. [DeVLor] p.61–62.

6. Exemples et exercices.

6.1.

- (1) Prendre $X = \mathcal{C}[a, b]$, $V =$ les constantes, et
 - (a) $\|\cdot\|_\infty$: on obtient $p = (\max f + \min f)/2$. Remarquer que $\|f - q\|_\infty$ a un *point critique non dérivable* en $q = p$.
 - (b) $\|\cdot\|_2$: $\|f - q\|_2^2$ est simplement un trinôme du second degré en q . On obtient $p = (b - a)^{-1} \int_a^b f(x) dx$.
 - (c) $\|\cdot\|_1$: supposer d'abord f monotone $\Rightarrow p = f((a + b)/2)$.
- (3) Prendre $X = \{f \in \mathcal{C}[0, 1] : f(0) = 0\}$ et $V = \{q(x) = ax\}_{a \in \mathbb{R}}$. Comparer avec Taylor $xf'(0)$ pour quelques fonctions simples de V , par exemple $\sin x$, etc.

- (4) **Systèmes surdéterminés d'équations linéaires.** On ne peut (normalement) pas résoudre exactement N équations linéaires à n inconnues si $N > n$. On choisit $\mathbf{x} \in \mathbb{R}^n$ tel que $\|\mathbf{b} - \mathbf{Ax}\|$ soit minimum selon une norme de \mathbb{R}^N . Cela correspond à $X = \mathbb{R}^N$ et $V =$ espace sous-tendu par les colonnes de \mathbf{A} .

$\|\cdot\|_2$: moindres carrés

Cf. G.A. Watson, *Approximation Theory and Numerical Methods*, Wiley, 1980, pour $\|\cdot\|_\infty$ et programmation linéaire: [Sti] § 2.7.

6.2. Moyenne et médiane.

Reprendre les constantes, mais pour approcher une fonction discrète, par exemple des cotes d'étudiants (exemple purement imaginaire⁴):

x	1	2	3	4	5
$f(x)$	18	17	16	14	2

Le bon sens suggère que l'examen a été bien réussi, qu'il n'y a qu'un cas malheureux à déplorer.

Pourtant, la meilleure constante selon la norme du maximum $\|\cdot\|_\infty$ est $(\min_i f(x_i) + \max_i f(x_i))/2 = 10$ suggère une assez médiocre performance d'ensemble: cette norme est trop sensible aux cas extrêmes pour être utile ici.

Selon $\|\cdot\|_2$, on obtient la *moyenne* $67/5 = 13.4$, encore curieusement basse: tous les étudiants sauf un seraient au-dessus de la moyenne...

La vraie performance moyenne (au demeurant excellente) est évidemment celle de l'étudiant classé en $(n/2)^{\text{ème}}$ position (*médiane*). Elle est obtenue en minimisant

$$\|f - c\|_1 = \sum_{i=1}^n |f(x_i) - c| \text{ selon } c, \text{ ce qui donne bien } c = f(x_{n/2}). \text{ (Plus précisément: } f(x_{(n+1)/2})$$

si n est impair; $f(x_{n/2})$ ou $f(x_{1+n/2})$ si n est pair: exemple de non unicité).

6.3. Principaux sous-espaces de fonctions utilisés en approximation.

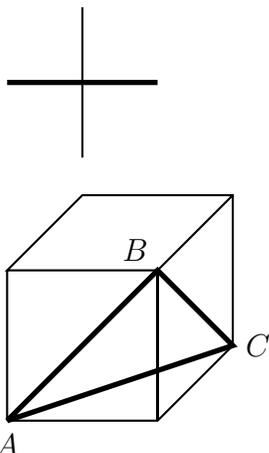
- (1) \mathcal{P}_n , polynômes de degré $\leq n$. Base usuelle: $\{1, x, x^2, \dots, x^n\}$. Dimension $n + 1$. On se préoccupera beaucoup de déterminer des bases plus efficaces...
- (2) \mathcal{T}_n , polynômes trigonométriques de degré $\leq n$. Bases les plus utilisées: $\{\{1, \cos x, \cos 2x, \dots, \cos nx\}, \{\sin x, \sin 2x, \dots, \sin nx\}\}$; $\{e^{-inx}, e^{-i(n-1)x}, \dots, e^{-ix}, 1, e^{ix}, \dots, e^{inx}\}$. Dimension $2n + 1$.
- (3) $\mathcal{S}_n^k(\Delta)$, fonctions spline définies sur $[a, b]$ à partir d'un découpage $\Delta = \{a = x_1 < x_2 < \dots < x_{N-1} < x_N = b\}$. $f \in \mathcal{S}_n^k(\Delta)$ si
 - (a) f est de classe \mathcal{C}^k dans $[a, b]$, et
 - (b) la restriction de f à un sous-intervalle $[x_i, x_{i+1}]$ est un polynôme de degré $\leq n$.
 Une base: $\{1, x, \dots, x^n, (x-x_2)_+^{k+1}, \dots, (x-x_2)_+^n, (x-x_3)_+^{k+1}, \dots, (x-x_3)_+^n, \dots, \dots, (x-x_{N-1})_+^{k+1}, \dots, (x-x_{N-1})_+^n\}$, où F_+ désigne la fonction qui vaut zéro quand $F(x) \leq 0$, et qui vaut $F(x)$ quand $F(x) \geq 0$. Dimension: $n + 1 + (n - k)(N - 2)$.

Base la plus élégante: B-splines, nulles sur le plus grand nombre possible de sous-intervalles.

- (4) \mathcal{E}_n^k , combinaisons de polynômes et d'exponentielles données: $\sum_{j=1}^k p_j(x)e^{\alpha_j x}, p_j \in \mathcal{P}_n$.

⁴d'après un exemple donné par J. Buijs, KULeuven.

6.4. Centre et rayon de Tchebycheff d'une partie P de X .



$c \in X$ est un centre de Tchebycheff de P si $\sup_{x \in P} \|c - x\|$ est la plus petite possible, cette norme minimale est appelée rayon de Tchebycheff de P . C'est donc le rayon de la plus petite boule contenant P . Par exemple⁵, tous les points du segment vertical $(0, -1) - (0, 1)$ sont des centres du segment horizontal $(-1, 0) - (1, 0)$ de \mathbb{R}^2 muni de $\|\cdot\|_\infty$.

Plus étonnant: où est (sont) le(s) centre(s) du triangle de sommets $A = (1, -1, -1)$, $B = (1, 1, 1)$ et $C = (-1, 1, -1)$ de \mathbb{R}^3 muni de $\|\cdot\|_\infty$? Le triangle est entièrement contenu dans la boule de rayon 1 centrée à l'origine (le cube de la figure ci-contre). Tout autre centre présumé (c_1, c_2, c_3) ne convient pas car au moins une des différences c_i – la $i^{\text{ème}}$ coordonnée d'un des points A, B , ou C a une valeur absolue > 1 (il y a toujours une de ces coordonnées qui vaut 1 et une qui vaut -1). La curiosité est donc que le centre n'est pas dans P . On montre que toute partie de X admet un centre dans son enveloppe convexe si et seulement si X est préhilbertien (voir ce mot en p. 76). Cf. Klee, Victor, Circumspheres and inner products. *Math. Scand.* **8** 1960 363–370. Garkavi, A. L. On the Čebyšev center and convex hull of a set. (Russian) *Uspehi Mat. Nauk* **19** 1964 no. 6 (120), 139–145. Holmes, Richard B. *A course on optimization and best approximation. Lecture Notes in Mathematics*, Vol. **257**. Springer-Verlag, Berlin-New York, 1972. viii+233 pp.

6.5. Largeurs de Kolmogorov. A un degré plus élevé de maturité, la théorie de l'approximation se pose le problème suivant: pour une norme donnée, et une partie P donnée de X , on considère *tous les sous-espaces V de dimension m de X* , et on évalue pour chaque V l'erreur du “plus mauvais” f de P : $M(V) = \sup_{f \in P} \min_{g \in V} \|f - g\|$. On minimise alors $M(V)$ selon V ... On ne sait donc pas à l'avance quelle sera la nature des approximations.

Exemple (A. Pinkus, *n-Widths in Approximation Theory*, Springer, 1985): $X = \{f : f \in \mathcal{C}^r[0, 2\pi] \text{ avec } f(0) = f(2\pi), f'(0) = f'(2\pi), \dots, f^{(r-1)}(0) = f^{(r-1)}(2\pi)\}$ muni de la norme de L^2 , $P = \{f : f \in X, \|f^{(r)}\| \leq 1\}$. On trouve $V = \mathcal{T}^{n-1}$, espace des polynômes trigonométriques de degré $\leq n - 1 =$ plus grand entier $< m/2$; $M(V)$ minimum $= 1/n^r$. C'est une façon de justifier l'intérêt numérique des *séries de Fourier*.

Cf. aussi C.A. Micchelli, T.J. Rivlin, Editors: *Optimal Estimation in Approximation Theory*, Plenum Press, New York, 1977.

6.6. Coapproximation.

“Soit E un espace vectoriel normé, G un sous-espace de E et $x \in E$. Tout élément $g_0 \in G$ vérifiant $\forall g \in G : \|g_0 - g\| \leq \|x - g\|$ est appelé élément de meilleure coapproximation de x par des éléments de G .”

Dans un préhilbertien, approximation et coapproximation sont équivalentes.

Cf. Papini, Pier Luigi; Singer, Ivan: Best coapproximation in normed linear spaces. *Monatsh. Math.* **88**, 27-44 (1979). Rao, Geetha S.; Swaminathan, M.: Best coapproximation and Schauder bases in Banach spaces. *Acta Sci. Math.* **54**, No.3/4, 339-354 (1990). Rao, Geeta S.; Muthukumar, S.: Semi-continuity properties of the coapproximation operator. *Math. Today* **5**, 37-48 (1987). Rao, Geetha S.; Chandrasekaran, K.R.: Characterizations of elements of best coapproximation in normed linear spaces. *Pure Appl. Math. Sci.* **26**, 139-147 (1987). Rao, Geetha S.; Swaminathan, M.: On normal bases. *Bull. Calcutta Math. Soc.* **88**, No.2, 107-112 (1996).

⁵Exemples et références fournis par Michel Coyette.

CHAPITRE 2

Approximation au sens de Tchebycheff.

1. Théorème d'équioscillation de Tchebycheff.

“Approximation au sens de Tchebycheff” signifie simplement meilleure approximation utilisant la norme du maximum $\| \cdot \|_\infty$, ce qui n'a rien de très banal, semble-t-il¹. La contribution cruciale de P.L. Tchebycheff (1821-1894) consista à décrire de façon saisissante cette meilleure approximation. Voyons immédiatement le théorème essentiel de Tchebycheff:

1.1. Théorème d'équioscillation de Tchebycheff (1853).²

Soit $X = \mathcal{C}[a, b]$ l'espace des fonctions continues réelles sur l'intervalle compact $[a, b]$, V l'espace \mathcal{P}_n des polynômes réels de degré $\leq n$. Alors \hat{p} est la meilleure approximation de f dans \mathcal{P}_n au sens de la norme du maximum si et seulement si on peut trouver $n + 2$ points distincts dans $[a, b]$

$$a \leq x_{n+1} < x_n < x_{n-1} < \dots < x_1 < x_0 \leq b$$

où $f - \hat{p}$ prend des valeurs de même valeur absolue $E = \|f - \hat{p}\|_\infty$ et de signes alternés:

$$f(x_i) - \hat{p}(x_i) = \sigma E (-1)^i, \quad i = 0, 1, \dots, n + 1, \quad (3)$$

où σ est le signe de $f(x_0) - \hat{p}(x_0)$. Un tel ensemble de points est appelé alternant .

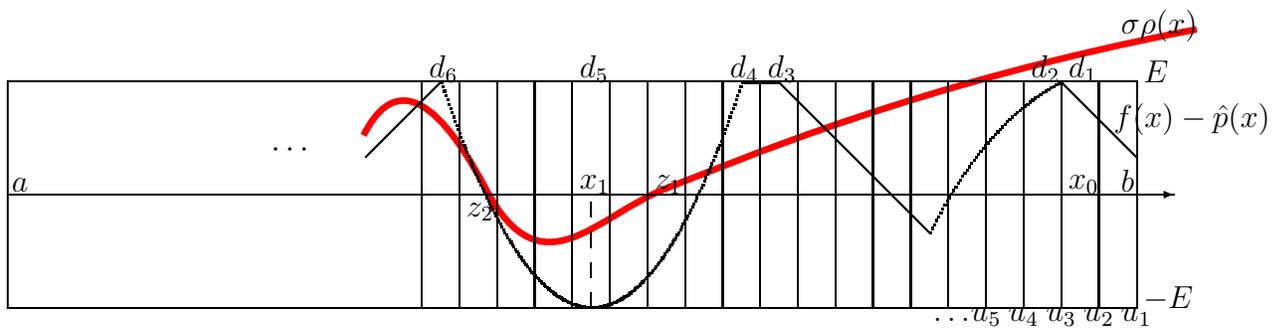
Il est en effet indispensable qu'il ne soit, en aucun endroit, susceptible de plus ou de moins. . . . Le point à partir duquel il est égal en tout sens tend également vers ses limites.

Parménide

On peut donc avoir plusieurs points, ou même tout un intervalle, où $f - \hat{p}$ atteint un extrémum, ces points ne comptent que pour un point tant que $f - \hat{p}$ n'atteint pas l'extrémum opposé (voir figure qui représente un exemple de graphe de $f - \hat{p}$).

¹On dit aussi “approximation minimax”.

² Très exactement, dans “Théorie des mécanismes connus sous le nom de parallélogrammes”, Mémoires présentés à l'Académie Impériale des sciences de St.-Petersbourg par divers savants, VII, 1854, pp. 539–568, *Lu le 28 janvier 1853*, où le théorème d'équioscillation est esquissé. Tchebycheff reprend la question de manière beaucoup plus détaillée dans un article de 1859 (lu le 9 octobre 1857) “Sur les questions de minima qui se rattachent à la représentation approximative des fonctions” [Tche]. Il a fallu attendre le 20^{ème} siècle et le développement de la notion de compacité (Borel) pour arriver à la démonstration complète [Che].



Les points on été numérotés à partir de la droite, on verra plus loin que c'est plus commode.

1.2. Preuve de la condition nécessaire: \hat{p} optimal dans $\mathcal{P}_n \Rightarrow (3)$.

1.2.1. *Deux premiers points.* On a donc $E = \|f - \hat{p}\| = \max_{a \leq x \leq b} |f(x) - \hat{p}(x)|$. La fonction continue $|f - \hat{p}|$ atteint en au moins un point de $[a, b]$ son maximum E . En fait, $f - \hat{p}$ atteint sa valeur maximale E_{\max} en au moins un point, et sa valeur minimale E_{\min} en au moins un autre point. Par définition de E , $E = \max(E_{\max}, -E_{\min})$.

En fait, $E = E_{\max} = -E_{\min}$: s'il en était autrement, par exemple si $\min_x (f(x) - \hat{p}(x)) = \mu > -E$, soit $\gamma := (E + \mu)/2$. Remarquons que $\gamma \neq 0$. Alors, $\hat{p} + \gamma$ serait meilleur que \hat{p} :

$$-(E - \gamma) = \mu - \gamma = \min_x (f(x) - \hat{p}(x) - \gamma) \leq \max_x (f(x) - \hat{p}(x) - \gamma) = E - \gamma,$$

on aurait donc $\|f - \hat{p} - \gamma\|_{\infty} = E - \gamma < E$.

Même raisonnement si on avait $\max_x (f(x) - \hat{p}(x)) < E$.

On a donc déjà un alternant d'au moins deux points. Prenons $n > 0$ puisqu'il n'y a plus rien à démontrer si $n = 0$.

1.2.2. *Découpage.* Comme $f - \hat{p}$ est continue sur le compact $[a, b]$, elle y est uniformément continue: $\forall \varepsilon > 0, \exists \delta > 0$ tel que dans *tout* intervalle de longueur δ dans $[a, b]$, l'oscillation de $f - \hat{p}$ est inférieure à ε . Prenons le δ correspondant à $\varepsilon = E/2$ et divisons $[a, b]$ en $s - 1$ intervalles³ $[u_s, u_{s-1}], [u_{s-1}, u_{s-2}], \dots, [u_2, u_1]$ de longueurs $\leq \delta$, avec $a = u_s < u_{s-1} < \dots < u_2 < u_1 = b$. Sur les intervalles $[u_{k+1}, u_k]$ où $|f - \hat{p}|$ atteint son maximum E , $|f - \hat{p}|$ ne peut prendre des valeurs inférieures à $E/2$, de sorte que $f - \hat{p}$ garde un signe constant sur ces intervalles-là (il y en a au moins deux). Notons, à partir de la droite, d_1, d_2, \dots, d_N ces intervalles. Sur les *autres* intervalles (fermés) $[u_{i+1}, u_i]$, la fonction continue $|f - \hat{p}|$ n'atteint jamais la valeur E , elle atteint donc une valeur maximale que nous notons E^* . On a $E^* < E$. Soit $\sigma = +1$ ou -1 le signe de $f - \hat{p}$ sur d_1 . Regroupons ces intervalles tant que $f - \hat{p}$ y garde le même signe:

$$\left. \begin{array}{ll} d_1, d_2, \dots, d_{k_1} & \text{signe} = \sigma \\ d_{k_1+1}, d_{k_1+2}, \dots, d_{k_2} & \text{signe} = -\sigma \\ \dots & \dots \\ d_{k_{m-1}+1}, d_{k_{m-1}+2}, \dots, d_{k_m} & \text{signe} = (-1)^{m-1} \sigma \end{array} \right\}$$

Dans le cas de la figure, on a $\sigma = 1, k_1 = 4, k_2 = 5, k_3 \geq 6 \dots$
 Montrons que $m \geq n + 2$.

³Ici, on suit [à peu près] les notations de Natanson [Nat] pp. 26 et suivantes; voir aussi [Riv1, chap. 1, § 1.2].

Supposons que $m < n + 2$. Comme la fonction continue $f - \hat{p}$ a le signe σ sur tout l'intervalle d_{k_1} et le signe opposé $-\sigma$ sur tout l'intervalle d_{k_1+1} , il y a au moins un intervalle $[u_{r+1}, u_r]$ entre d_{k_1+1} et d_{k_1} . Soit $z_1 \in (u_{r+1}, u_r)$ (l'intervalle ouvert!). On peut d'ailleurs prendre un point entre d_{k_1+1} et d_{k_1} où $f - \hat{p}$ s'annule. De même, on choisit z_2 entre d_{k_2+1} et d_{k_2} , ..., z_{m-1} entre $d_{k_{m-1}+1}$ et $d_{k_{m-1}}$.

1.2.3. *Construction d'un polynôme meilleur que \hat{p} si $m < n + 2$.* . On construit alors

$$\rho(x) = (x - z_1)(x - z_2) \dots (x - z_{m-1}).$$

C'est un polynôme de degré $m - 1 < n + 1$, donc un élément de $V = \mathcal{P}_n$. Le polynôme ρ ne s'annule qu'aux z_i , donc garde un signe constant sur chaque intervalle d_i . Le polynôme $\sigma\rho$ a précisément le même signe que $f - \hat{p}$ sur chaque d_i : sur d_1, \dots, d_{k_1} , $\sigma\rho(x)$ et $f(x) - \hat{p}(x)$ ont tous deux le signe σ puisque $x > z_1 > z_2 > \dots : \rho(x) > 0$; sur $d_{k_1+1}, \dots, d_{k_2}$, $\sigma\rho(x)$ et $f(x) - \hat{p}(x)$ ont tous deux le signe $-\sigma$ puisque $z_1 > x > z_2 > \dots : \rho(x) < 0$, etc. Montrons que l'on peut trouver $\lambda > 0$ assez petit tel que $\hat{p} + \lambda\sigma\rho$ soit meilleur que \hat{p} : sur les d_i , $f(x) - \hat{p}(x) - \lambda\sigma\rho(x)$ garde le signe de $f(x) - \hat{p}(x)$ pourvu que $\lambda|\rho(x)|$ soit toujours plus petit que $|f(x) - \hat{p}(x)| \geq E/2$, il suffit donc de prendre $0 < \lambda < E/(2\|\rho\|_\infty)$. Le maximum de $|f - \hat{p} - \lambda\sigma\rho|$ sur les d_i sera donc réduit à un nombre inférieur à $E - \lambda \min_{x \in \cup d_i} |\rho(x)| < E$. Enfin, sur les intervalles $[u_{i+1}, u_i]$ qui ne sont pas des d_i , on sait que $|f(x) - \hat{p}(x)|$ ne dépasse pas $E^* < E$, donc, avec $\lambda < (E - E^*)/\|\rho\|$, le maximum de $|f - \hat{p} - \lambda\sigma\rho|$ reste encore inférieur à $\max_{x \notin \cup d_i} |f - \hat{p}| + \lambda\|\rho\| < E$.

Si $m < n + 2$, on peut donc construire une approximation meilleure que \hat{p} dans \mathcal{P}_n , ce qui est impossible: $m \geq n + 2$. \square

1.3. Preuve de la condition suffisante (3) $\Rightarrow \hat{p}$ optimal dans \mathcal{P}_n .

Montrons que, si on a pu construire $\hat{p} \in \mathcal{P}_n$ tel que $f - \hat{p}$ possède un alternant de $n + 2$ points, \hat{p} ne peut être amélioré: si $q \in \mathcal{P}_n$ était meilleure approximation que \hat{p} , on aurait $\|f - q\|_\infty < E = \|f - \hat{p}\|_\infty$ (on ne considère que le cas $f \notin \mathcal{P}_n : E > 0$). Examinons ce qui se passe en x_0, x_1, \dots, x_{n+1} : $f(x_0) - \hat{p}(x_0) = \sigma E$ et $|f(x_0) - q(x_0)| < E$, donc, $q(x_0) - \hat{p}(x_0) = f(x_0) - \hat{p}(x_0) - (f(x_0) - q(x_0))$ est non nul avec le signe σ . En x_1 , $q - \hat{p}$ a le signe $-\sigma$, etc. Le polynôme $q - \hat{p}$ devrait donc s'annuler en $n + 1$ points distincts séparant les $n + 2$ points x_i , ce qui n'est pas possible avec un polynôme non nul de degré $\leq n$. \square

1.4. **Exemple.** *La fonction $f(x) = \sin 6\pi x$ ne peut être approchée sur $[0, 1]$ par un polynôme meilleur que le polynôme nul tant que le degré reste ≤ 4 .*
En effet, avec $p(x) \equiv 0$, $f - p$ présente un alternant de 6 points en $1/12, 3/12, \dots, 11/12$, de sorte que 0 est optimal dans $\mathcal{P}_0, \dots, \mathcal{P}_4$.

1.5. **Théorème d'unicité de la meilleure approximation polynomiale au sens de Tchebycheff.** *La meilleure approximation polynomiale de degré $\leq n$ au sens de Tchebycheff d'une fonction réelle continue sur un compact $[a, b]$ est unique.*

Comme la norme de Tchebycheff n'est pas stricte, on ne peut appliquer un théorème général, mais le raisonnement particulier suivant établit la proposition:

Si p et q sont deux polynômes de meilleure approximation, il en est de même de $(p + q)/2$ (combinaison convexe). Donc, par le théorème d'équioscillation, $f - (p + q)/2$ possède un alternant de $n + 2$ points $\{x_0, x_1, \dots, x_{n+1}\}$ (condition nécessaire). En un de ces points, soit x_k , on a $f(x_k) - (p(x_k) + q(x_k))/2 = \sigma_k E$, où $\sigma_k = (-1)^k \sigma$ est le signe approprié, donc $f(x_k) - p(x_k) + f(x_k) - q(x_k) = 2\sigma_k E$. Comme $|f(x_k) - p(x_k)|$ et $|f(x_k) - q(x_k)|$ sont tous

deux $\leq E$, cela n'est possible que si $f(x_k) - p(x_k) = \sigma_k E$ et $f(x_k) - q(x_k) = \sigma_k E$, donc $q(x_k) = p(x_k)$ en $n+2$ points distincts, le polynôme $q-p$ devrait s'annuler en ces $n+2$ points, impossible avec un polynôme non nul de degré $\leq n$. \square

2. Propriétés de la meilleure approximation.

2.1. Symétrie. Théorème. *Si f est une fonction paire sur un intervalle de la forme $[-a, a]$, c'est-à-dire si $\forall x \in [-a, a], f(-x) = f(x)$, alors sa meilleure approximation \hat{p} de degré $\leq n$ au sens de Tchebycheff est également une fonction paire. De même, si f est une fonction impaire ($f(-x) = -f(x)$), \hat{p} est une fonction impaire⁴.*

En effet, soit $f(-x) = sf(x)$ avec $s = 1$ (fonction paire) ou $s = -1$ (fonction impaire). On a

$$E = \|f(x) - \hat{p}(x)\|_\infty = \|f(-x) - \hat{p}(-x)\|_\infty = \|sf(-x) - s\hat{p}(-x)\|_\infty = \|f(x) - s\hat{p}(-x)\|_\infty,$$

d'où on conclut que $s\hat{p}(-x)$ est également une meilleure approximation de f sur le même intervalle $[-a, a]$, d'où, par unicité, $s\hat{p}(-x) = \hat{p}(x)$. \square

Si f est paire sur $[-a, a]$, la meilleure approximation de degré $\leq 2n$ présente donc nécessairement un alternant d'au moins $2n+3$ points (puisque le même \hat{p} est également optimal dans \mathcal{P}_{2n+1})!

Par exemple, $x^2 + 1/8$ est la meilleure approximation de degré ≤ 3 de $|x|$ sur $[-1, 1]$. Les points extrémaux de la fonction d'erreur sont bien a nombre de 5: $0, \pm 1/2$ et ± 1 .

2.2. Théorème (de La Vallée Poussin). *Soit f continue sur le compact $[a, b]$ et \hat{p} sa meilleure approximation de degré $\leq n$ au sens de Tchebycheff. Soit $E = \|f - \hat{p}\|_\infty$. Alors, si on connaît $p \in \mathcal{P}_n$ et des points*

$$a \leq x'_{n+1} < x'_n < \dots < x'_1 < x'_0 \leq b$$

où $f - p$ prend des valeurs de signes alternés, on a

$$\min_{0 \leq i \leq n+1} |f(x'_i) - p(x'_i)| \leq E.$$

En effet, si tous les $|f(x'_i) - p(x'_i)|$ étaient $> E$, chaque $\hat{p}(x'_i) - p(x'_i)$ serait du même signe que $f(x'_i) - p(x'_i)$, puisque $\hat{p}(x'_i) - p(x'_i) = f(x'_i) - p(x'_i) - (f(x'_i) - \hat{p}(x'_i))$ et que $f(x'_i) - \hat{p}(x'_i)$ n'est pas assez grand (sa valeur absolue est $\leq E$) pour renverser le signe de $f(x'_i) - p(x'_i)$.

Le polynôme $\hat{p} - p \in \mathcal{P}_n$ devrait donc prendre des signes alternés en $n+2$ points, donc changer de signe en $n+1$ points intermédiaires, ce qui est impossible⁵. \square

Ce dernier théorème permet d'apprécier dans quelle mesure un polynôme $p \in \mathcal{P}_n$ est proche de \hat{p} : si p est tel que les signes des $f(x'_i) - p(x'_i)$ soient alternés, on a l'encadrement pour E

$$\min_i |f(x'_i) - p(x'_i)| \leq E \leq \max_{a \leq x \leq b} |f(x) - p(x)|,$$

puisque p n'est (normalement) pas optimal et que le dernier terme n'est autre que $\|f - p\|_\infty$.

Si les deux normes $E = \|f - \hat{p}\|$ et $\|f - p\|$ sont proches, p et \hat{p} sont également proches, par forte unicité (Chap. 1, § 4.2, p. 25, qu'on démontrera dans un instant): $\|p - \hat{p}\| \leq (\|f - p\| - E)/\gamma$.

⁴Cas particulier d'un théorème d'invariance, [Mei] pp. 25-26.

⁵**Exercice.** Montrez cette variante: soit un polynôme $p \in \mathcal{P}_n$ tel que $f - p$ a des signes alternés en $n+2$ points $a \leq x'_{n+1} < x'_n < \dots < x'_0 \leq b$, alors, pour *tout* polynôme $q \in \mathcal{P}_n$, on a, en au moins un des points x'_i , $|f(x'_i) - p(x'_i)| \leq |f(x'_i) - q(x'_i)|$. On en déduit évidemment $\min_i |f(x'_i) - p(x'_i)| \leq \|f - q\|_\infty$. (J. Hendrickx, MAP 22, 2002-2003.) (En effet, s'il n'en était pas ainsi, $q - p = f - p - (f - q)$ présenterait des signes alternés en les $n+2$ points x'_j).

2.3. Unicité forte. [Che] essayons d'apprécier dans quelle mesure $\|f - p\|$ est plus grand que $\|f - \hat{p}\|$ quand $p \neq \hat{p}$. En chacun des $n + 2$ points de l'alternant de $f - \hat{p}$, on a

$$f(x_i) - p(x_i) = f(x_i) - \hat{p}(x_i) + \hat{p}(x_i) - p(x_i) = (-1)^i \sigma \left(E + \|\hat{p} - p\| \frac{Q(x_i)}{(-1)^i \sigma} \right),$$

où Q est le polynôme de norme unité $(\hat{p} - p)/\|\hat{p} - p\| \in \mathcal{P}_n$. Les $n + 2$ valeurs $Q(x_i)/((-1)^i \sigma)$ ne peuvent être toutes ≤ 0 , sinon Q aurait au moins $n + 1$ zéros. On montre que $\max_i (-1)^i \sigma Q(x_i)$ est une fonction strictement positive, continue en les coefficients de Q sur le compact $B_1 \cap \mathcal{P}_n$, elle admet donc un minimum également strictement positif que l'on note γ .

En un x_i où $(-1)^i \sigma Q(x_i) \geq \gamma$, on a donc $|f(x_i) - p(x_i)| \geq E + \|\hat{p} - p\| \gamma$, d'où $\|f - p\| \geq E + \gamma \|\hat{p} - p\|$. \square

2.4. Signes alternés. A partir d'une approximation quelconque p , il n'est pas toujours possible de trouver $n + 2$ points x'_i où $f - p$ prend des valeurs de signes alternés, mais on peut se donner $n + 2$ points x'_i arbitraires (toujours avec $a \leq x'_{n+1} < \dots < x'_0 \leq b$, bien sûr), et construire $p \in \mathcal{P}_n$ tel que les valeurs $f(x'_i) - p(x'_i)$ soient de signes alternés.

Propriété de Haar, espaces de Haar.

Les fonctions continues $\Phi_1, \Phi_2, \dots, \Phi_m$ de A vers \mathbb{R} ou \mathbb{C} ont la **propriété de Haar** si

- (1) A contient au moins m points,
- (2) toute combinaison linéaire $\sum_1^m a_i \Phi_i$ est, soit la fonction nulle sur A , soit a au plus $m - 1$ zéros distincts dans A .

L'espace \mathcal{H}_m sous-tendu par $\Phi_1, \Phi_2, \dots, \Phi_m$ est alors appelé **espace de Haar** ([DeVLor] pp.67–73, [Mei] § 4.1).

Exemples.

- (1) \mathcal{P}_n , avec $m = n + 1$;
- (2) l'espace des polynômes pondérés $p(x) = w(x)q(x)$, $q \in \mathcal{P}_n$ sur un ensemble A où w ne s'annule pas; ceci permet de traiter des problèmes de meilleure erreur **relative** avec $w = 1/f$ si f ne s'annule pas sur A ;
- (3) l'espace \mathcal{T}_n des polynômes trigonométriques $p(\theta) = a_0 + \sum_1^n (a_k \cos(k\theta) + b_k \sin(k\theta))$ sur $A = [-\pi, \pi)$: $p(\theta) = e^{-in\theta} q(e^{i\theta})$ avec $q \in \mathcal{P}_{2n}$, on a donc $m = 2n + 1$;
- (4) les combinaisons linéaires réelles d'exponentielles réelles $f(x) = \sum_1^m a_k e^{\alpha_k x}$, sur un intervalle réel A , avec $\alpha_1, \dots, \alpha_m$ distincts: la propriété est vraie pour $m = 1$, si elle est vraie pour $m - 1$, f ne peut avoir m zéros réels distincts, car il en serait de même pour g : $g(x) = e^{-\alpha_m x} f(x)$, donc g' aurait encore $m - 1$ zéros réels distincts dans A (th. de Rolle), mais g' est une combinaison de $m - 1$ exponentielles.

Soit $\Phi_1, \Phi_2, \dots, \Phi_m$ une suite libre, donc une base de \mathcal{H}_m . Si \mathcal{H}_m est un espace de Haar, et si x_1, x_2, \dots, x_m sont des points distincts de A , le déterminant

$$D(x_1, x_2, \dots, x_m) = \begin{vmatrix} \Phi_1(x_1) & \cdots & \Phi_m(x_1) \\ \vdots & & \vdots \\ \Phi_1(x_m) & \cdots & \Phi_m(x_m) \end{vmatrix} \quad (4)$$

est non nul: si on veut exprimer qu'une combinaison linéaire $\sum_1^m a_j \Phi_j$ s'annule en $x = x_1, x_2, \dots, x_m$, on aboutit au système d'équations linéaires homogènes $\sum_1^m \Phi_j(x_i) a_j = 0$, $i = 1, \dots, m$, qui ne peut admettre que la solution triviale $a_1 = \dots = a_m = 0$, donc le déterminant, qui n'est autre que (4), doit être non nul.

Passons maintenant au cas de fonctions réelles sur un intervalle réel A . Soit $x_1 > x_2 > \dots > x_m$. $D(x_1, x_2, \dots, x_m)$ est alors de signe constant, soit σ : en effet, $D(x_1, x_2, \dots, x_m)$ est une fonction réelle continue de chaque x_i qui ne peut s'annuler tant que x_1, x_2, \dots, x_m sont distincts⁶. Ensuite: $\text{sign } D(x, x_1, x_2, \dots, x_{m-1}) = \sigma(-1)^k$ si $x_{k+1} < x < x_k$, en effet, on permute la première et la $k^{\text{ème}}$ ligne pour retrouver un déterminant de signe σ .

⁶Par homotopie: soit $\{x_1^{(0)}, \dots, x_m^{(0)}\}$, avec $x_1^{(0)} > x_2^{(0)} > \dots > x_m^{(0)}$ une configuration de départ, et soit σ le signe de $D(x_1^{(0)}, \dots, x_m^{(0)})$, alors $D((1-t)x_1^{(0)} + tx_1, \dots, (1-t)x_m^{(0)} + tx_m)$ est une fonction réelle continue de t qui ne peut pas s'annuler sur $t \in [0, 1]$, doit donc garder le signe σ .

On voit maintenant comment construire p dans un espace de Haar réel \mathcal{H}_m de base Φ_1, \dots, Φ_m sur $A \subset \mathbb{R}$, tel que $f - p$ prenne des valeurs de même valeur absolue et de signes alternés en $m + 1$ points donnés $x_1 > x_2 > \dots > x_{m+1}$ de A : les $m + 1$ équations sont

$$f(x_i) - p(x_i) = f(x_i) - \sum_{j=1}^m \Phi_j(x_i) a_j = e(-1)^{i-1}, \quad i = 1, 2, \dots, m + 1,$$

où les $m + 1$ inconnues sont a_1, \dots, a_m et e . Le système est donc

$$\begin{bmatrix} \Phi_1(x_1) & \cdots & \Phi_m(x_1) & 1 \\ \Phi_1(x_2) & \cdots & \Phi_m(x_2) & -1 \\ \Phi_1(x_3) & \cdots & \Phi_m(x_3) & 1 \\ \vdots & & \vdots & \vdots \\ \Phi_1(x_{m+1}) & \cdots & \Phi_m(x_{m+1}) & (-1)^m \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \\ e \end{bmatrix} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \\ \vdots \\ f(x_{m+1}) \end{bmatrix}$$

de déterminant

$$\sum_{i=1}^{m+1} (-1)^m D(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{m+1}),$$

somme de termes non nuls tous de même signe $\sigma(-1)^m$, donc non nul.

2.5. Algorithme d'échange. (Remez⁷)

A partir d'un alternant d'essai $a \leq x'_{n+1} < x'_n < \dots < x'_0 \leq b$, on détermine $p \in \mathcal{P}_n$ tel que les $n + 2$ valeurs $f(x'_i) - p(x'_i)$, $i = 0, \dots, n + 1$ soient de même valeur absolue et de signes alternés, par la méthode vue plus haut:

$$f(x'_i) - p(x'_i) = (-1)^i e', \quad i = 0, \dots, n + 1.$$

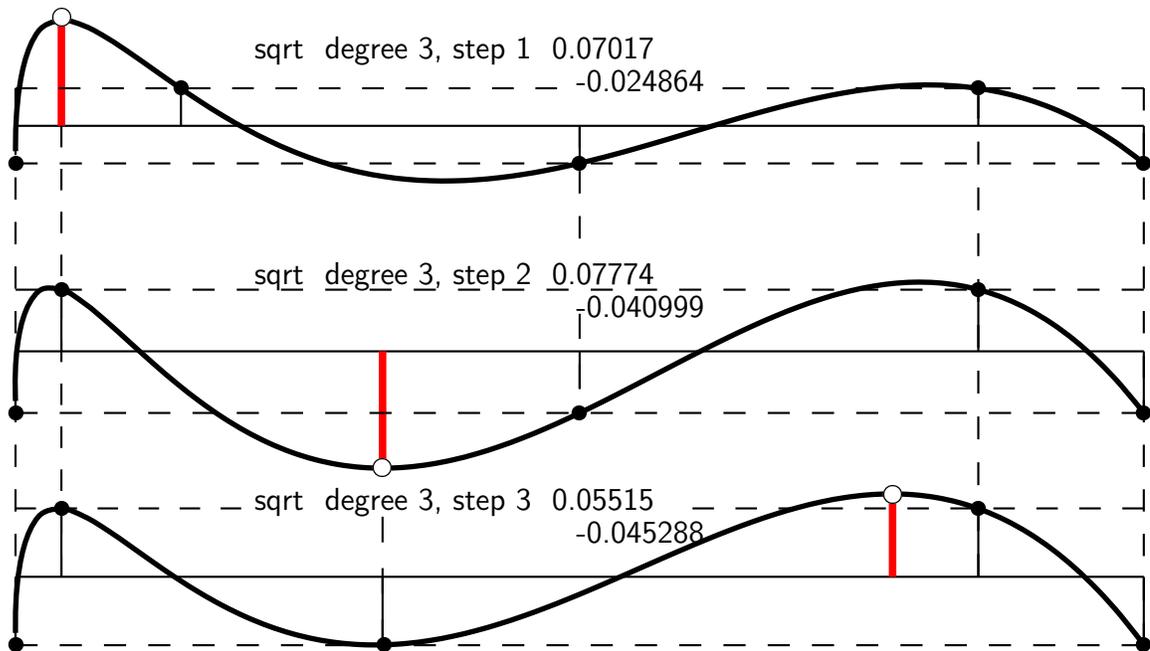
On examine alors $f(x) - p(x)$ sur tout $[a, b]$ (ou sur une grille fine où les valeurs de f sont données). Soit x'' un point de $[a, b]$ où $|f - p|$ atteint son maximum E' . Par le théorème de La Vallée Poussin et la définition de E , norme de l'erreur de meilleure approximation, on a

$$|e'| \leq E \leq E'.$$

On va maintenant construire un nouvel alternant d'essai contenant x'' et $n + 1$ des points de l'ancien alternant d'essai, en veillant à ce que les signes de $f - p$ soient encore alternés sur le nouvel alternant. Ainsi,

- Si x'' est situé entre deux points x'_{j+1} et x'_j , on rejette celui de ces deux points où $f - p$ a le même signe que $f(x'') - p(x'')$,
- Si $a \leq x'' < x'_{n+1}$, on rejette x'_{n+1} si $f - p$ y a le même signe qu'en x'' ; sinon, le nouvel alternant se compose de $x'', x'_{n+1}, x'_n, \dots, x'_1$ (donc, on rejette x'_0).
- Si $x'_0 < x'' \leq b$, on rejette x'_0 si $f - p$ y a le même signe qu'en x'' ; sinon, le nouvel alternant se compose de $x'_n, \dots, x'_1, x'_0, x''$ (donc, on rejette x'_{n+1}).

⁷Yevgeny Yakovlevich Remez, 17 février 1896– 31 août 1975, cf. *J. Approx. Theory* **84** (1996) pp. 119–120.



Etapes de l'algorithme d'échange.

On calcule maintenant $q \in \mathcal{P}_n$ tel que les valeurs de $f - q$ soient de même valeur absolue $|e''|$ et de signes alternés sur le nouvel alternant. Ce polynôme q est la meilleure approximation possible de f sur le nouvel alternant (équioscillation), donc, en appliquant le théorème de La Vallée Poussin sur le nouvel alternant:

$$|e'| = \min_{x \in \text{nouvel alternant}} |f(x) - p(x)| \leq |e''| \leq E.$$

On obtient ainsi une suite *croissante* $|e'| \leq |e''| \leq \dots$ bornée supérieurement par E , donc convergente (vers E).

La figure ci-dessus montre un exemple d'application (\sqrt{x} sur $[0, 1]$, $n = 3$). Chaque configuration est précédée des valeurs de E' et e' correspondantes. On est parti d'un alternant d'essai donnant $\dots |e'| = 0.025$, nettement plus petite que $E' = \|f - p_{\text{essai}}\|_{\infty} = 0.070$. Puis, les échanges successifs rapprochent ces deux bornes, jusqu'à arriver à $E = 0.045929$.

Exercice: essayer $|x|$ sur $[-1, 1]$, degré 6... [remezp.htm](#)

Variantes (échanges multiples), vitesse de convergence: cf. [Pow, chap. 8 & 9].

Programmes matlab : [remezp.m](#); java: [remezp.htm](#); javascript: [remezjs.htm](#)

2.6. Meilleure approximation au sens $\| \cdot \|_{\infty}$ sur un compact quelconque.

Soit f une fonction continue à valeurs réelles ou complexes définie sur le compact quelconque A , et V un espace vectoriel de dimension finie de fonctions continues sur A . $p \in V$ est alors une meilleure approximation de f dans V si et seulement si on a, pour $\forall q \in V$,

$$\max_{x \in A_0} \operatorname{Re} \left\{ [f(x) - p(x)] \overline{q(x)} \right\} \geq 0,$$

où A_0 est l'ensemble des points de A où $|f(x) - p(x)| = \|f - p\|_{\infty}$. (Théorème de Kolmogorov, [DeVLor] pp.63–67, [Mei] § 3.1).

Algorithmes dans le plan complexe:

<http://www.math.mu-luebeck.de/workers/modersitzki/COCA/coca5.html>

3. Polynômes de Tchebycheff.

3.1. Meilleure approximation d'un polynôme de degré n dans \mathcal{P}_{n-1} .

La fonction continue la plus simple qui ne soit *pas* dans $V = \mathcal{P}_{n-1}$ est un polynôme $f(x) = \alpha x^n + \dots$ de degré n . La fonction d'erreur $\hat{e} := f - \hat{p}$ est encore un polynôme de degré n ayant en commun avec f son coefficient de x^n . Les deux problèmes suivants sont équivalents:

- (1) Construire la meilleure approximation \hat{p} de $f(x) = \alpha x^n + \alpha' x^{n-1} + \dots$ dans \mathcal{P}_{n-1} sur $[a, b]$,
- (2) Construire β', β'', \dots tels que $\hat{e}(x) = \alpha x^n + \beta' x^{n-1} + \beta'' x^{n-2} + \dots$ soit de norme $\|\cdot\|_\infty$ minimale sur $[a, b]$.

En effet, si on résout 1), il suffit de prendre $\hat{e} = f - \hat{p}$; si on résout 2), $\hat{p} = f - \hat{e}$. La formulation 2) montre qu'il suffit de s'occuper de α, a et b . De plus, si on résout 1) ou 2) avec $\alpha = 1$, il suffira de multiplier la réponse \hat{p} ou \hat{e} par α . On peut donc se limiter au problème:

- (3) Construire la meilleure approximation de x^n dans \mathcal{P}_{n-1} sur $[a, b]$.

Enfin, on se ramène à l'intervalle canonique $[-1, 1]$ par le changement de variable:

$$t \text{ parcourt } [-1, 1] \iff x = \frac{a+b}{2} + \frac{b-a}{2} t \text{ parcourt } [a, b]. \quad (5)$$

On peut donc se limiter à:

- (4) Construire la meilleure approximation de x^n dans \mathcal{P}_{n-1} sur $[-1, 1]$.

3.2. Définition. Pour $n \geq 1$, le **polynôme de Tchebycheff** de degré n est

$$T_n(x) = \frac{\hat{e}_n(x)}{E_n} = \frac{x^n - \hat{p}_{n-1}(x)}{E_n}, \quad (6)$$

où \hat{p}_{n-1} est la meilleure approximation au sens de Tchebycheff de x^n sur $[-1, 1]$, et $E_n = \|\hat{e}_n\|_\infty$ est la norme de l'erreur de meilleure approximation; pour $n = 0$, on définit $T_0(x) \equiv 1$.

3.3. Premières propriétés. Comme \hat{e}_n est une fonction d'erreur de meilleure approximation dans \mathcal{P}_{n-1} , elle atteint les valeurs E_n et $-E_n$ en au moins $n - 1 + 2 = n + 1$ points distincts de $[-1, 1]$, par le théorème d'équioscillation:

$$-1 \leq x_n < x_{n-1} < \dots < x_1 < x_0 \leq 1$$

en prenant des signes alternés: $\hat{e}_n(x_k) = \sigma(-1)^k E_n$, $k = 0, \dots, n$, ou encore $T_n(x_k) = \sigma(-1)^k$, $k = 0, \dots, n$. Le polynôme T_n de degré n doit donc s'annuler en un point de chaque intervalle (ouvert) (x_{k+1}, x_k) , $k = 0, \dots, n - 1$, ce qui fait déjà n zéros, il ne peut y en avoir plus:

- (1) Si $n \geq 1$, les zéros de T_n sont tous réels et distincts dans $(-1, 1)$.

Soient z_0, z_1, \dots, z_{n-1} ces n zéros.

Entre deux zéros de T_n , on est assuré de trouver au moins un zéro de la dérivée T'_n (théorème de Rolle), mais T'_n est de degré $n - 1$ et a donc $n - 1$ zéros au plus:

- (2) Si $n \geq 2$, les zéros de la dérivée T'_n sont tous réels et distincts dans $(-1, 1)$.

Ces $n - 1$ points de $(-1, 1)$ sont des extrema possibles de T_n . Mais comme il faut $n + 1$ extrema dans $[-1, 1]$, on doit bien prendre ces $n - 1$ points intérieurs et leur adjoindre les deux extrémités -1 et 1 :

$$T_n(1) = \sigma ; T_n(-1) = \sigma(-1)^n ; T_n(x_k) = \sigma(-1)^k \text{ et } T'_n(x_k) = 0, k = 1, \dots, n - 1. \quad (7)$$

Remarquons que $\sigma = 1$: le polynôme de degré n $x^n - \hat{p}_{n-1}(x)$ a un coefficient unité de x^n et n zéros z_0, \dots, z_{n-1} , donc $x^n - \hat{p}_{n-1}(x) = \prod_{k=0}^{n-1} (x - z_k)$, qui est strictement positif en $x = 1$ puisque tous les $z_k < 1$. Et comme T_n a le même signe que $x^n - \hat{p}_{n-1}(x)$ (E_n étant > 0 dans (6)), $T_n(1) > 0$.

Exploitions maintenant (7): $1 - T_n^2$ est un polynôme ≥ 0 (puisque $\|T_n\|_\infty = 1 \Rightarrow -1 \leq T_n(x) \leq 1$) dans $[-1, 1]$, qui s'annule en $-1, 1$, et en x_1, \dots, x_{n-1} . En ces $n - 1$ derniers points intérieurs, $1 - T_n^2$ a des zéros de multiplicité paire, car $1 - T_n^2$ ne devient < 0 nulle part dans $[-1, 1]$. Comme le degré de $1 - T_n^2$ vaut $2n$, on ne peut avoir que des zéros intérieurs *double*:

$$1 - T_n(x)^2 = \frac{1}{E_n^2} (1 - x^2) \prod_{k=1}^{n-1} (x - x_k)^2, \quad (8)$$

où la constante $1/E_n^2$ assure l'égalité des coefficients de x^{2n} dans les deux membres de (8) (d'après (6), le coefficient de x^n de $T_n(x)$ est $1/E_n$).

Exprimons maintenant que les x_k intérieurs ($k = 1, \dots, n - 1$) sont les zéros de la dérivée T'_n (de coefficient de x^{n-1} égal à n/E_n):

$$T'_n(x) = \frac{n}{E_n} \prod_{k=1}^{n-1} (x - x_k)$$

$$(T'_n(x))^2 = \frac{n^2}{E_n^2} \prod_{k=1}^{n-1} (x - x_k)^2,$$

et comparons avec (8)!

$$(1 - x^2) (T'_n(x))^2 = n^2 (1 - T_n(x)^2). \quad (9)$$

Cette relation (9) est une *équation différentielle* dont T_n est une solution. On va en tirer une première *formule explicite* de T_n :

(3) **Proposition.** *Pour toute détermination de la fonction arccos, on a*

$$\boxed{T_n(x) = \cos(n \arccos x)} \quad -1 \leq x \leq 1, n = 0, 1, \dots \quad (10)$$

ce qui s'exprime aussi par la représentation paramétrique

$$\boxed{x = \cos \theta, T_n(x) = \cos(n\theta)} \quad \theta \text{ réel}, n = 0, 1, \dots \quad (11)$$

En effet, soit $y = T_n$. L'équation (9) $(1 - x^2)y'^2 = n^2(1 - y^2)$ est singulière en $x = \pm 1$ et ne permet pas à elle seule de déterminer $y = T_n$. Mais nous savons que $y(1) = 1$, que y est un polynôme croissant entre x_1 (encore inconnu) et $x_0 = 1$, donc,

(a) si $x_1 < x < x_0 = 1$, $y'(x) > 0$, $-1 < y(x) < 1$, d'où

$$\frac{y'}{\sqrt{1 - y^2}} = \frac{n}{\sqrt{1 - x^2}},$$

toutes les quantités présentes sont positives. On intègre aisément:

$$- \arccos y = C - n \arccos x,$$

soit $y = \cos(n \arccos x - C)$. En $x = 1$, toute détermination de arccos est un multiple entier de 2π , et on doit avoir $y = 1$, donc C est un multiple entier arbitraire de 2π et la solution est bien (10). Remarquons que l'on peut donner une formulation correcte, mais moins élégante, en termes de la fonction arcsin. Enfin, x_1 est la première abscisse < 1 (en parcourant $(-1, 1)$ de droite à gauche) où $y = -1$. On trouve $x_1 = \cos(\pi/n)$.

(b) Si $x_2 < x < x_1 = \cos(\pi/n)$, $y'(x) < 0$, $-1 < y(x) < 1$, d'où

$$\frac{y'}{\sqrt{1-y^2}} = -\frac{n}{\sqrt{1-x^2}},$$

qui s'intègre maintenant en $-\arccos y = C' + n \arccos x$, ou encore $y = \cos(n \arccos x - C')$ (parité du cosinus!). C'est encore la forme (10), et la condition limite $y(x_1) = -1$ donne encore $C' = 0$ (à un multiple entier de 2π près, mais ça ne change rien à y).

(c) Les autres intervalles $[x_{k+1}, x_k]$ reproduisent la situation des deux premiers. \square
Nous pouvons maintenant préciser la position des extrema et des zéros:

(4) Les $n + 1$ extrema de T_n sur $[-1, 1]$ sont

$$x_k = \cos(k\pi/n), \quad k = 0, 1, \dots, n.$$

(5) Les n zéros de T_n sont

$$z_k = \cos((k + 1/2)\pi/n), \quad k = 0, 1, \dots, n - 1.$$

En principe, nous pouvons maintenant évaluer $E_n = \|\hat{e}_n\|_\infty$ par $E_n = |$ valeur de $\hat{e}_n|$ en n'importe quel extremum x_m :

$$\begin{aligned} E_n &= \left| \prod_{k=0}^{n-1} (x_m - z_k) \right| = \prod_{k=0}^{n-1} |\cos(m\pi/n) - \cos((k + 1/2)\pi/n)| \\ &= 2^n \prod_{p=1}^{2n} |\sin((p - 1/2)\pi/(2n))| = 2^n \prod_{p=1}^{2n} \left| \frac{1 - e^{i(p-1/2)\pi/n}}{2} \right| = 2^{1-n}, \end{aligned}$$

les $e^{i(p-1/2)\pi/n}$ étant les racines $2n^{\text{èmes}}$ de -1 . (cf. un théorème de Cotes-Wessel).

Exercice. Examiner θ complexe: $\theta = \theta_r + i\theta_i$, alors, $x = \cos \theta = (e^{i\theta} + e^{-i\theta})/2 = (e^{-\theta_i} e^{i\theta_r} + e^{\theta_i} e^{-i\theta_r})/2 = \cosh \theta_i \cos \theta_r - i \sinh \theta_i \sin \theta_r$; $T_n(x) = \cosh n\theta_i \cos n\theta_r - i \sinh n\theta_i \sin n\theta_r$.

3.4. Premiers échantillons. Un peu de virtuosité en formules trigonométriques permet d'obtenir quelques premiers polynômes de Tchebycheff.

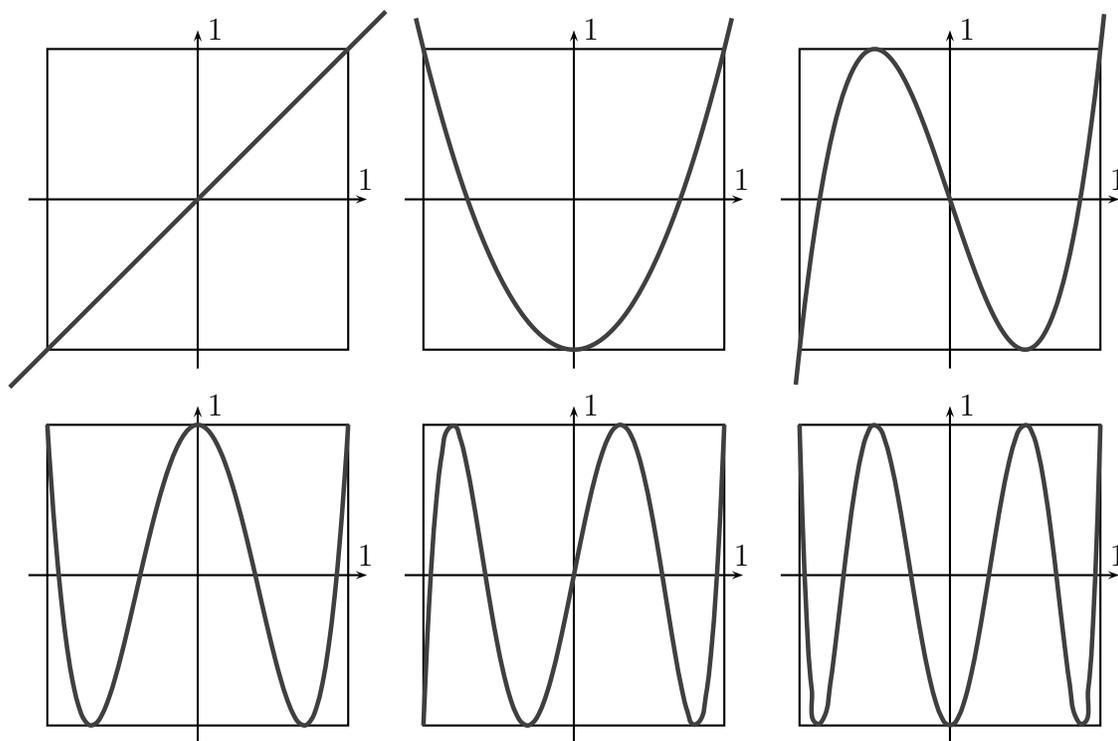
$T_0(x) = 1$. Par convention (compatible avec la formule (10)).

$T_1(x) = x$.

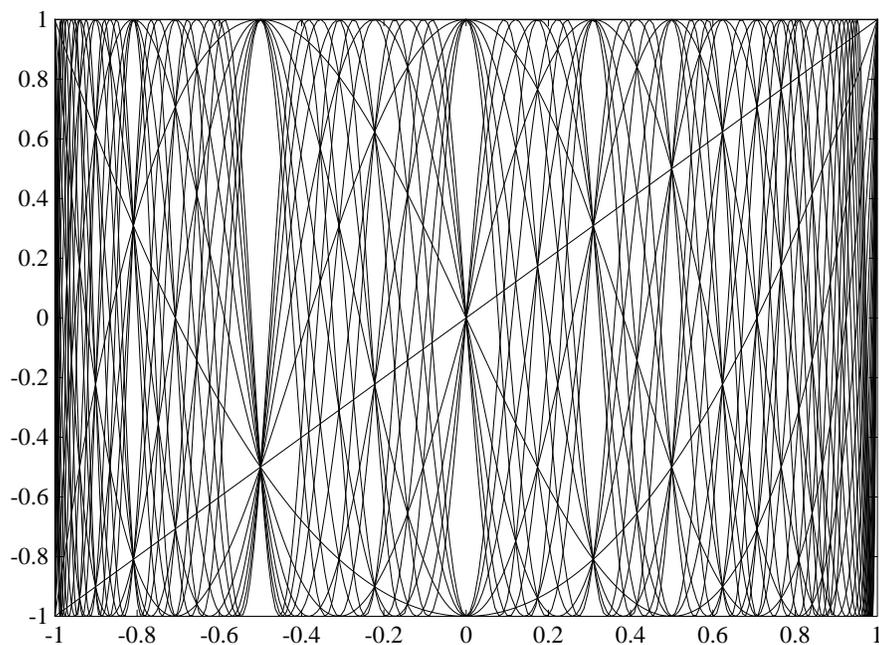
$\cos 2\theta = \cos^2 \theta - \sin^2 \theta = 2 \cos^2 \theta - 1$: $T_2(x) = 2x^2 - 1$.

$\cos 3\theta = \cos 2\theta \cos \theta - \sin 2\theta \sin \theta = (2 \cos^2 \theta - 1) \cos \theta - 2(1 - \cos^2 \theta) \sin \theta$: $T_3(x) = 4x^3 - 3x$.

$\cos 4\theta = \cos(2(2\theta)) = 2 \cos^2(2\theta) - 1 = 2(2 \cos^2 \theta - 1)^2 - 1$: $T_4(x) = 8x^4 - 8x^2 + 1$.

Graphes de T_1 à T_6 .

La figure suivante représente une superposition de quelques dizaines de polynômes de Tchebycheff.



Elle fut conçue et réalisée, dans l'enthousiasme que l'on devine, par Y. Brandsteert et A. Ninane, alors (1980) étudiants en physique. Leur ouvrage ne manqua la possibilité d'une publication que de peu, suite à la parution d'un article⁸ sur ce phénomène intéressant.

⁸E.L. Ortiz, T.J. Rivlin, Another look at the Chebyshev polynomials, *The American Math. Monthly* **90** (1983) 3-10.

3.5. Exercice. Montrez que $T_n(0)$ ne peut prendre que les valeurs $-1, 0$ et 1 ; que $T_n(-1/2)$ ne peut prendre que les deux valeurs $-1/2$ et 1 .

Utilisez (11).

Quelles sont les solutions de $T_m(x) = T_n(x)$? ($m \neq n$).

On peut maintenant expliquer quelques propriétés de la figure ci-dessus...

$T_m(x) = T_n(x) \Rightarrow \cos m\theta = \cos n\theta \Rightarrow (m \pm n)\theta = 2k\pi$. Les abscisses des points d'intersection sont donc des cosinus de multiples rationnels de π , et il en est de même des ordonnées $\cos m\theta$.

Voir aussi dans <http://icm.mcs.kent.edu/reports/1999/chebpol.pdf>,

<http://icm.mcs.kent.edu/reports/1998/cheby1.pdf> des relations, par M.O. Rayes, entre polynômes de Tchebycheff et nombres premiers, discussion dans CHEBPRIM.HTM

Figures de Lissajous⁹. La courbe (11) $(x, y = T_n(x)) : (\cos \theta, \cos(n\theta))$ est un cas particulier de figure de Lissajous. Ces figures permettent d'étudier des phénomènes ondulatoires. Voir un programme java dans <http://robmorton.20m.com/chaos/lissajou/lissajou.htm>

3.6. Relation de récurrence. La relation suivante est des plus importantes: on a

$$\boxed{T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)}, \quad n = 1, 2, \dots \quad (12)$$

En effet,

$$\begin{aligned} T_{n+1}(x) &= \cos(n+1)\theta = \cos \theta \cos n\theta - \sin \theta \sin n\theta \\ T_{n-1}(x) &= \cos(n-1)\theta = \cos \theta \cos n\theta + \sin \theta \sin n\theta \\ T_{n+1}(x) + T_{n-1}(x) &= 2 \cos \theta \cos n\theta = 2xT_n(x) \end{aligned}$$

où on a bien sûr utilisé la représentation (11). □

La récurrence (12) livre aussi la façon la plus simple de se persuader que T_n est bien un polynôme de degré n : à partir des conditions initiales (en n) $T_0(x) = 1$ et $T_1(x) = x$, on voit bien que T_{n+1} est un polynôme dont le degré vaut une unité de plus que le degré de T_n . De plus,

Proposition. Si $n \geq 1$, le coefficient de x^n de $T_n(x)$ vaut 2^{n-1} .

En effet, d'après (12), le coefficient de x^{n+1} de $T_{n+1}(x)$ est le *double* du coefficient de x^n de $T_n(x)$, puisque T_{n-1} n'est que de degré $n-1$ et n'a pas de coefficient de x^{n+1} . Le coefficient de x^n de $T_n(x)$ est donc constante fois 2^n , déterminé par le cas $n=1, T_1(x) = x$: la constante vaut $1/2$. □

On peut enfin préciser maintenant la norme d'erreur de la meilleure approximation de x^n dans \mathcal{P}_{n-1} :

Proposition. Si $n \geq 1$, la norme E_n d'erreur de la meilleure approximation de x^n par un polynôme de degré $\leq n-1$ vaut $E_n = \|\hat{e}_n\|_\infty = \|x^n - \hat{p}_{n-1}(x)\|_\infty = \frac{1}{2^{n-1}}$.

En effet, par la définition (6) de T_n quand $n \geq 1$, $T_n(x) = (x^n - \hat{p}_{n-1}(x))/E_n$, donc le coefficient de x^n de T_n , que nous savons valoir 2^{n-1} , est l'inverse de E_n . □

On peut maintenant donner complètement la solution du problème donné en 3.1, 1 :

Remarque. La meilleure approximation d'un polynôme de degré n $f(x) = \alpha x^n + \alpha' x^{n-1} + \dots$ dans \mathcal{P}_{n-1} sur $[a, b]$ est \hat{p}_{n-1} tel que

$$f(x) - \hat{p}_{n-1}(x) = 2\alpha \left(\frac{b-a}{4} \right)^n T_n \left(\frac{2x-a-b}{b-a} \right). \quad (13)$$

La norme d'erreur vaut donc $2|\alpha|[(b-a)/4]^n$.

⁹Remarque faite par L. Demanet, printemps 2000.

En effet, on se ramène d'abord à un problème sur $[-1, 1]$ par (5): $f(x) = \alpha x^n + \dots = \alpha[(a+b)/2 + t(b-a)/2]^n + \dots$, où les termes négligés ne contiennent que des puissances $\leq n-1$ de t . f est donc un polynôme de degré n en t , de coefficient de t^n égal à $\beta := \alpha[(b-a)/2]^n$. La meilleure approximation sur $t \in [-1, 1]$ donnera la fonction d'erreur $(\beta/2^{n-1})T_n(t)$, et on revient à la variable x par $t = [x - (a+b)/2]/[(b-a)/2]$. \square

3.7. Polynôme de moindre norme sur un intervalle, sous contrainte $p(0) = 1$.

Le problème se rencontre dans l'étude de méthodes itératives (Flanders-Shortley): trouver $p \in \mathcal{P}_n$ de norme minimale sur $[a, b]$, avec $p(0) = 1$ (on suppose $0 < a < b$). Solution:

$$p(x) = \frac{T_n\left(\frac{2x - a - b}{b - a}\right)}{T_n((-a - b)/(b - a))}.$$

Démonstration directe: si $q \in \mathcal{P}_n$ était meilleur que p , $p(x) - q(x)$ prendrait des valeurs de signes alternés aux $n + 1$ points de l'alternant de p sur $[a, b] \Rightarrow p - q$ devrait s'annuler en au moins n points de (a, b) ET en 0... \square

3.8. Meilleure approximation d'un polynôme de degré n dans \mathcal{P}_{n-2} . Ne s'obtient (généralement) pas en passant de n à $n-1$, puis à $n-2$! Le polynôme de la forme $x^n + \beta x^{n-1} + \dots$, avec β imposé, de moindre norme est de la forme constante $T_n(\gamma x + \delta)$ si $|\beta| \leq n \operatorname{tg}^2 \frac{\pi}{2n}$, sinon, c'est beaucoup plus compliqué (premier problème de Zolotarev, [Ach] appendice E, J. Todd, Applications of transformation theory: a legacy from Zolotarev (1847–1878), pp. 207–245 in *Approximation Theory and Spline Functions*, S.P. Singh & al., eds., Reidel, 1984, *Math. Intelligencer* 10 (1988) 50–53.).

3.9. Meilleure approximation de fonction rationnelle. cf. [Ach] avec $f(x) = 1/(x-a)$, $a > 1$, on trouve

$$f(x) - \hat{p}_n(x) = \frac{4\alpha^{n+2}}{(1-\alpha^2)^2} \left(e^{in\theta} \frac{\alpha - e^{i\theta}}{1 - \alpha e^{i\theta}} + e^{-in\theta} \frac{1 - \alpha e^{i\theta}}{\alpha - e^{i\theta}} \right),$$

où $\alpha = a - \sqrt{a^2 - 1} < 1$.

Exercice. La meilleure approximation de degré $\leq n$ de $f(x) = (1 + a^2 x^2)^{-1}$ en erreur *relative* est l'interpolant de f aux zéros de T_{2m} , où $m = \lfloor n/2 \rfloor + 1$.

Cf. R. Freund: On some approximation problems for complex polynomials, *Constr. Approx.* 4 (1988) 111-121, D.S.Lubinsky: Best approximation and interpolation of $(1 + (ax)^2)^{-1}$ and its transforms, *J. Approx. Th.* 125 (2003) 106-115. En effet, $p(x) = \frac{1 - C_m T_{2m}(x)}{1 - a^2 x^2}$, où C_m est tel que le numérateur soit divisible par

le dénominateur, a bien un degré $= 2m - 2 \leq n$, vérifie $\frac{f(x) - p(x)}{f(x)} = C_m T_{2m}(x)$ qui présente un alternant de $2m + 1 \geq n + 2$ points dans $[-1, 1]$.

3.10. Fonctions rationnelles de moindre et plus grande déviation. Trouver r rationnelle, aussi petite que possible sur $[-\alpha, \alpha]$ par rapport à ses valeurs sur $(-\infty, -\beta] \cup [\beta, \infty)$, c'est-à-dire qui minimise le rapport $\frac{\max_{|x| \leq \alpha} |r(x)|}{\min_{|x| \geq \beta} |r(x)|}$ (α et β , $0 < \alpha < \beta$ donnés). Troisième problème de Zolotarev ([Ach], Todd, *op. cit.*). On trouve une équation différentielle de la forme

$$\frac{r'^2}{(E^2 - r^2)(r^2 - F^2)} = \frac{c}{(x^2 - \alpha^2)(\beta^2 - x^2)}.$$

Par un changement de variable, on obtient une représentation paramétrique $r(x) = \Phi(u)$, $x = \Psi(u)$, où Φ et Ψ sont des solutions d'équations différentielles de la forme $Y' = \sqrt{P(Y)}$, avec P polynôme de degré 3 ou 4 (*fonctions elliptiques*).

Applications: filtres, méthodes itératives (directions alternées).

3.11. Propriétés extrémales des polynômes de Tchebycheff.

On va examiner ici des énoncés assurant que le polynôme de Tchebycheff T_n maximise telle ou telle propriété parmi tous les polynômes de \mathcal{P}_n de *norme unité*. La propriété en question sera d'abord la valeur absolue d'une *forme* sur \mathcal{P}_n , et il s'agira d'évaluer la *norme d'une forme*, aussi paraît-il utile de rappeler l'équivalence suivante:

Proposition. *Soit λ une forme (c'est-à-dire une application linéaire vers \mathbb{R} ou \mathbb{C}) sur un espace vectoriel normé V de dimension finie. Alors, si λ n'est pas la forme nulle sur V ,*

$$\|\lambda\| = \max_{\substack{f \in V \\ f \neq 0}} \frac{|\lambda(f)|}{\|f\|} = \frac{1}{\min_{\substack{g \in V \\ \lambda(g)=1}} \|g\|} = \frac{1}{\min_{\substack{g_0 \in V \\ \lambda(g_0)=0}} \|g_1 - g_0\|}, \quad (14)$$

où g_1 est un élément particulier de V vérifiant $\lambda(g_1) = 1$.

(Si V est de dimension infinie, il faut remplacer les 'max' et 'min' par des 'sup' et 'inf').

En effet, $|\lambda(f)|/\|f\| = |\lambda(Cf)|/\|Cf\|$ pour tout scalaire $C \neq 0$. On passe d'une égalité à l'autre en prenant $g = Cf$ avec $C = 1/\lambda(f)$. Tous les g possibles parcourent un espace affine, et on peut les représenter par $g = g_1 - g_0$, avec $g_0 \in V_0 = V \cap \text{Ker}(\lambda)$. \square

La forme en $1/\min \|g_1 - g_0\|$ est évidemment un problème de meilleure approximation dans l'espace vectoriel V_0 . On doit donc estimer $\min \|g_1 - g_0\|$ sur tous les $g_0 \in V_0$, erreur de meilleure approximation de g_1 dans V_0 .

Voici quelques-uns de ces énoncés:

- (1) *Parmi tous les polynômes réels de \mathcal{P}_n de norme $\|\cdot\|_\infty$ unité sur $[-1, 1]$, le polynôme de Tchebycheff T_n a le plus grand coefficient de x^n .*

En effet, soit $f(x) = \sum_0^n a_k(f)x^k$, ici, $\lambda(f) = a_n(f)$. On a $g(x) = f(x)/a_n(f) = x^n + \dots$, $g_1(x) = x^n$ (par exemple), $V_0 = \mathcal{P}_{n-1}$. Donc, par la définition (6) (p. 35) des polynômes de Tchebycheff, $g = g_1 - g_0$ de norme minimale = $E_n T_n$, $|a_n(f)|/\|f\|_\infty = 1/E_n = a_n(T_n)$ qui vaut, rappelons-le, 2^{n-1} (si $n \geq 1$). \square

- (2) *Parmi tous les polynômes réels de \mathcal{P}_n de norme $\|\cdot\|_\infty$ unité sur $[-1, 1]$, le polynôme de Tchebycheff T_n a le plus grand coefficient de x^k si k et n ont la même parité; T_{n-1} a le plus grand coefficient de x^k si k et n sont de parités différentes.*

Bien sûr, on a déjà traité $k = n$ au point précédent; $k = 0$ est assez trivial; on ne démontrera ici que le cas $k = n - 1$:

il faut montrer que $\hat{g}(x) = T_{n-1}(x)/a_{n-1}(T_{n-1}) = x^{n-1} + \dots$ est le polynôme de \mathcal{P}_n de moindre norme $\|\cdot\|_\infty$ sur $[-1, 1]$ parmi les polynômes ayant un coefficient de x^{n-1} égal à l'unité. Supposons que η soit un tel polynôme, c'est-à-dire $\eta(x) = a_n(\eta)x^n + x^{n-1} + a_{n-2}x^{n-2} + \dots$ de norme strictement inférieure à celle de \hat{g} , et examinons $\hat{g} - \eta$ aux n extrema $x_k = \cos(k\pi/(n-1))$, $k = 0, \dots, n-1$ de \hat{g} . En x_k , $\hat{g} - \eta$ a le même signe $(-1)^k$ que \hat{g} (comme chaque fois que l'on a comparé un polynôme à un polynôme de Tchebycheff), donc $\hat{g} - \eta$ a au moins $n-1$ zéros z_1, \dots, z_{n-1} dans $(-1, 1)$, avec $x_k < z_k < x_{k-1}$, $k = 1, \dots, n-1$. $\hat{g} - \eta$ est un polynôme de \mathcal{P}_n sans coefficient de x^{n-1} . Si le coefficient de x^n était lui-même nul, $\hat{g} - \eta$ ne serait que de degré $\leq n-2$ et ne pourrait admettre $n-1$ zéros, donc $\hat{g} - \eta$ est de degré exact n , admet déjà $n-1$ zéros réels, donc un $n^{\text{ème}}$ zéro réel z_n . Où se trouve z_n ? La somme des zéros d'un polynôme de degré n vaut l'opposé du rapport du coefficient de x^{n-1} et de celui, non nul, de x^n , ici, 0:

$$-1 = x_n = -x_0 - \dots - x_{n-1} < z_n = -z_1 - \dots - z_{n-1} < -x_1 - \dots - x_n = x_0 = 1.$$

Tous les n zéros de $\hat{g} - \eta$ sont donc dans $(-1, 1)$, mais alors le signe de $\hat{g}(-1) - \eta(-1)$ vaut $(-1)^{n+1}$ et non $(-1)^n$, contradiction, ouf! \square

Bel exemple de démonstration *ad hoc*.

Le cas des coefficients de Taylor en $x = 1$ est plus facile à établir:

- (3) Parmi tous les polynômes réels de \mathcal{P}_n de norme $\| \cdot \|_\infty$ unité sur $[-1, 1]$, le polynôme de Tchebycheff T_n a les plus grands coefficients de Taylor en $x = 1$.

En effet, soit $f(x) = \sum_{j=0}^n b_j(f)(x-1)^j$ le développement de Taylor autour de 1 d'un élément $f \in \mathcal{P}_n$. Il faut montrer que $\hat{g} = T_n/b_k(T_n)$ est le polynôme de \mathcal{P}_n , ayant son coefficient de Taylor b_k fixé à 1, de moindre norme. Supposons que $\eta \in \mathcal{P}_n$, avec $b_k(\eta) = 1$ ait une norme strictement plus petite que celle de \hat{g} . On examine alors (air connu) la différence $\hat{g} - \eta$ aux $n + 1$ extrema $x_\ell = \cos(\ell\pi/n)$, $\ell = 0, \dots, n$ de \hat{g} , et on trouve, bien entendu, que $\hat{g}(x_\ell) - \eta(x_\ell)$ a le même signe $(-1)^\ell$ que $\hat{g}(x_\ell)$. $\hat{g} - \eta$ a donc tous ses n zéros dans $(-1, 1)$. Appliquons k fois le théorème de Rolle: la dérivée $k^{\text{ème}}$ $\hat{g}^{(k)} - \eta^{(k)}$ a tous ses $n - k$ zéros dans $(-1, 1)$, mais alors

$$0 = b_k(\hat{g} - \eta) = \frac{\hat{g}^{(k)}(1) - \eta^{(k)}(1)}{k!}$$

ne peut pas être nul: contradiction. □

Au fait, les valeurs extrémales de ces coefficients sont

$$b_0(T_n) = 1, b_1(T_n) = n^2, b_k(T_n) = 2^k \frac{n^2(n^2 - 1)(n^2 - 4) \cdots (n^2 - (k - 1)^2)}{(2k)!}$$

(dériver $k - 1$ fois (21) et porter $x = 1$).

Ces coefficients sont très élevés quand n est grand. Comme tous les $b_k(T_n)$ sont positifs, on obtient leur somme en évaluant $T_n(x) = \sum_{k=0}^n b_k(T_n)(x-1)^k$ en $x = 2$, et $T_n(2) > (2 + \sqrt{3})^n / 2$, le plus grand des $b_k(T_n)$ est donc supérieur à $(2 + \sqrt{3})^n / (2(n+1))$.

Valeur en un point fixé $\notin (-1, 1)$:

- (4) Parmi tous les polynômes réels de \mathcal{P}_n de norme $\| \cdot \|_\infty$ unité sur $[-1, 1]$, le polynôme de Tchebycheff T_n prend la plus grande valeur absolue en tout point réel fixé $\xi \notin (-1, 1)$. Montrons que $\hat{g}(x) = T_n(x)/T_n(\xi)$ a la plus petite norme possible parmi tous les polynômes qui prennent la valeur 1 en ξ : sinon, si η était un tel polynôme de \mathcal{P}_n de norme strictement inférieure à celle de \hat{g} , $\hat{g}(x_k) - \eta(x_k)$ prendrait le même signe que $\hat{g}(x_k) = \text{constante} \times T_n(x_k)$ aux $n + 1$ extrema $x_k = \cos(k\pi/n)$, $k = 0, \dots, n$ de T_n , donc $\hat{g} - \eta$ aurait n zéros dans $(-1, 1)$, mais ce polynôme de \mathcal{P}_n doit déjà s'annuler en $\xi \notin (-1, 1)$. □

Ce raisonnement a déjà été utilisé dans le § 3.7, p. 40.

Le polynôme de Tchebycheff est encore extrémal avec une valeur ponctuelle d'une dérivée d'ordre quelconque:

$$f \text{ réel} \in \mathcal{P}_n, \xi \text{ réel} \notin (-1, 1) \Rightarrow |f^{(k)}(\xi)| \leq |T_n^{(k)}(\xi)| \|f\|_\infty, \quad k = 0, 1, \dots, n.$$

cf. [Riv2, § 2.7].

Bien sûr, il n'y a pas lieu de discuter des valeurs dans $(-1, 1)$, fatalement bornées par $\| \cdot \|_\infty$, mais on peut examiner les plus grandes valeurs prises par des dérivées d'un polynôme dans $[-1, 1]$. Remarquons que l'on a déjà rencontré des dérivées en 0 et en 1 avec les coefficients a_k et b_k . Mais il s'agit maintenant de normes d'opérateurs.

- (5) **Inégalités de Bernstein et des Markoff**

- (a) Si g est un polynôme trigonométrique de degré $\leq n$ (c'est-à-dire une combinaison linéaire de $1, \sin \theta, \cos \theta, \sin 2\theta, \cos 2\theta, \dots, \sin n\theta, \cos n\theta$: $g \in \mathcal{T}_n$),

$$\left\| \frac{dg}{d\theta} \right\|_{\infty} \leq n \|g\|_{\infty}$$

(*inégalité de Bernstein*).

- (b) $f \in \mathcal{P}_n \implies \|f'\|_{\infty} \leq T'_n(1) \|f\|_{\infty} = n^2 \|f\|_{\infty}$ (A.A. Markoff, 1890).
 (c) $f \in \mathcal{P}_n \implies \|f^{(k)}\|_{\infty} \leq T_n^{(k)}(1) \|f\|_{\infty}$
 $= \frac{n^2(n^2-1)(n^2-4)\cdots(n^2-(k-1)^2)}{1 \cdot 3 \cdot 5 \cdots (2k-1)} \|f\|_{\infty}$, $k = 1, 2, \dots, n$ (W.A.

Markoff).

- (d) Les inégalités précédentes restent valables si on remplace $\|f\|_{\infty}$ par $\max_{0 \leq j \leq n} |f(\cos(j\pi/n))|$ (Duffin & Schaeffer, 1941).

(i) Preuve de Bernstein [Mha, III.3 B]:

(A) inégalité de Stechkin: $g \in \mathcal{T}_n \rightarrow \|dg/d\theta\|_{\infty} \leq (n/2)\omega_g(\pi/n)$.

(B) Lemme de M. Riesz (1915): si $h \in \mathcal{T}_n$ est extrémal en θ_0 , $|h(\theta)| \geq \|h\|_{\infty} \cos(n[\theta - \theta_0])$ dans $|\theta - \theta_0| < \pi/(2n)$.

Preuve du lemme: sinon, la fonction $\text{sign } h(\theta_0)h(\theta) - \cos(n[\theta - \theta_0])\|h\|$, qui a déjà un zéro double en θ_0 , aurait au moins un zéro de plus dans l'intervalle de longueur π/n centré en θ_0 , et doit encore s'annuler dans les $2n-1$ autres intervalles de longueur π/n constituant une période complète (h est partout inférieur à sa norme). Cela fait $2n+1$ zéros sur une période, impossible (propriété de Haar p. 32).

(C) Preuve de Stechkin: soit $h = g'$. Donc, g' ne change pas de signe dans $(\theta_0 - \pi/(2n), \theta_0 + \pi/(2n))$, et l'intégrale de g' sur cet intervalle, qui est une différence de deux valeurs de g , donc $\leq \omega_g(\pi/n)$, a une valeur absolue supérieure à l'intégrale de $\|g'\|$ fois $\cos(n[\theta - \theta_0])$, ce qui est bien $(2/n)\|g'\|$.

(D) Enfin, $\forall u, \omega_g(u) \leq 2\|g\|$. □

(ii) Preuve de Markoff.

Pour une démonstration moderne complète de l'inégalité de W.A. Markoff, voir [Riv2] § 2.7. Voici une façon d'arriver à prouver l'inégalité de A.A. Markoff et d'autres relations intéressantes (J. Todd, *Introduction to the Constructive Theory of Functions*, Birkhäuser (ISNM 1), 1963, chap. 4):

(A) $f \in \mathcal{P}_{n-1} \implies \|f\|_{\infty} \leq n \max_{-1 \leq x \leq 1} (1-x^2)^{1/2} |f(x)|$.
 Démonstration par interpolation aux zéros de T_n .

(B) $f \in \mathcal{P}_n \implies |f'(x)| \leq n \|f\|_{\infty} / (1-x^2)^{1/2}$, $-1 < x < 1$.
 Conséquence de l'inégalité de Bernstein avec $g(\theta) = f(\cos \theta)$.

(C) On établit enfin l'inégalité de A.A. Markoff en appliquant le point ci-dessus à f' :
 $\|f'\|_{\infty} \leq n \max(1-x^2)^{1/2} |f'(x)| \leq n \times n \|f\|_{\infty}$. □

Cf. aussi [DeVLor] pp. 97-104

- (e) *Inégalité de Remez*: $\forall f \in \mathcal{P}_n$, si $\mu(f)$ est la somme des longueurs des intervalles où $|f(x)| \leq \delta$ dans $[-1, 1]$, alors $\|f\|_{\infty} \leq \delta T_n \left(\frac{4}{\mu(f)} - 1 \right)$. ([Fr], pp. 119-121).

3.12. Autres propriétés des T_n .

(1) T_n est une fonction paire ou impaire selon la parité de n .

En effet, $f(x) = x^n$ est une fonction paire ou impaire sur $[-1, 1]$ selon la parité de n . D'après 2.1 p. 31, il en est donc de même pour \hat{p}_{n-1} , $\hat{e}_n = f - \hat{p}_{n-1}$, donc pour $T_n = \hat{e}_n/E_n$ d'après la définition (6).

Vérification directe d'après (10):

$$T_n(-x) = \cos[n \arccos(-x)] = \cos[n(\pi + \arccos x)] = (-1)^n \cos[n \arccos x] = (-1)^n T_n(x). \quad \square$$

(2) $T_m(T_n(x)) = T_{mn}(x)$.

En effet, les deux membres valent $\cos mn\theta$, d'après (11). □

Problème. Montrez que f_n de degré exact n , $n = 0, 1, \dots$, $f_n \circ f_m = f_m \circ f_n$, $m, n = 0, 1, \dots \Rightarrow f_n(x) = c + a^{n-1}(x-c)^n$ ou $f_n(x) = c + aT_n((x-c)/a)$, $n = 0, 1, \dots$

En effet, soit $f_0 = \alpha \Rightarrow$ tous les $f_n(\alpha) = \alpha$. Si β est la deuxième racine de $f_2(x) = \alpha$, on a $f_2(x) = \alpha + \xi(x-\alpha)(x-\beta) = \alpha - \xi(\alpha-\beta)^2/4 + \xi(x-c)^2$, où $c = (\alpha + \beta)/2$. $f_n \circ f_2 = f_2 \circ f_n$ et f_n de degré exact $n \Rightarrow f_n - c$ est de parité n en $x - c$. $f_3(x) = x + \xi^2(x-\alpha)(x-c)(x-\beta)$ et

on trouve $\xi(\alpha - \beta) = 2$ ou 4 . Il reste alors $[n/2]$ inconnues dans $f_n(x) = c + \sum_0^{[n/2]} \gamma_k(x-c)^{n-2k}$,

avec $\gamma_0 = \xi^{n-1}$. On trouve alors de proche en proche $\gamma_1, \gamma_2, \dots$ en égalant les coefficients de $(x-c)^{2n-2}, (x-c)^{2n-4}, \dots$ dans $f_n(f_2(x)) \equiv f_2(f_n(x))$. Il n'y a donc que deux familles possibles, que l'on identifie alors aisément (!).

En particulier, les *polynômes de Tchebycheff sur $[0, 1]$* sont

$$T_n^*(x) := T_n(2x - 1) = T_{2n}(\sqrt{x}).$$

(3) On a l'expression algébrique explicite

$$T_n(x) = \frac{(x + \sqrt{x^2 - 1})^n + (x - \sqrt{x^2 - 1})^n}{2}. \quad (15)$$

On établit cette formule au moyen de la notion de **solution de récurrence**:

Définition. Une suite $\mathbf{u} = \{u_n\}_{n=0}^\infty$ vérifie une **relation de récurrence** (ou simplement **récurrence**, ou encore **équation aux différences**) d'ordre p si ses éléments vérifient des équations de la forme suivante:

$$F_n(u_n, u_{n+1}, \dots, u_{n+p}) = 0, \quad n = 0, 1, \dots$$

L'expression "équation aux différences" se justifie par la possibilité que l'on a toujours de réécrire une récurrence comme une relation entre $u_n, \Delta u_n, \dots, \Delta^p u_n$, où $\Delta u_n := u_{n+1} - u_n$, $\Delta^2 u_n := \Delta u_{n+1} - \Delta u_n, \dots$, $\Delta^p u_n := \Delta^{p-1} u_{n+1} - \Delta^{p-1} u_n$. Comme pour les équations différentielles, la solution d'une récurrence d'ordre p est normalement déterminée par p conditions, par exemple par la détermination des p conditions initiales u_0, u_1, \dots, u_{p-1} .

Proposition. L'ensemble des solutions de la récurrence **linéaire homogène** d'ordre p

$$u_{n+p} = a_n u_{n+p-1} + b_n u_{n+p-2} + \dots + z_n u_n, \quad n = 0, 1, \dots \quad (16)$$

est un espace vectoriel de dimension p .

En effet, on voit que toute combinaison linéaire $\{\alpha u_n + \beta v_n\}_0^\infty$ de deux solutions de (16) est encore une solution, et que les solutions $\mathbf{u}^{(k)}$, $k = 1, \dots, p$ déterminées par: $[u_0^{(k)}, u_1^{(k)}, \dots, u_{p-1}^{(k)}] = k^{\text{ème}}$ ligne de la matrice unité d'ordre p constituent une base

de l'ensemble des solutions. Toute solution est \mathbf{u} donnée par $\mathbf{u} = u_0\mathbf{u}^{(1)} + u_1\mathbf{u}^{(2)} + \dots + u_{p-1}\mathbf{u}^{(p)}$

Changement de base: les p solutions $\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(p)}$ forment encore une base si le déterminant des $v_i^{(j)}$, $i = 0, 1, \dots, p-1$, $j = 1, \dots, p$ est non nul.

Une base remarquable: soit ρ_1 et ρ_2 les deux racines de $\rho^2 - a\rho - b = 0$. La récurrence d'ordre 2 à coefficients constants $u_{n+2} = au_{n+1} + bu_n$ admet la base $\{\{\rho_1^n\}, \{\rho_2^n\}\}$ si $\rho_1 \neq \rho_2$; $\{\{\rho_1^n\}, \{n\rho_1^n\}\}$ si $\rho_1 = \rho_2$.

Intéressantes considérations sur la suite de *Fibonacci* dans le cours de X. Hubaut <http://bib3.ulb.ac.be/coursmath/fibo.html>

Démontrons maintenant (15): on a $a = 2x$ et $b = -1$. La récurrence (12) admet comme solutions des combinaisons des puissances $n^{\text{èmes}}$ de ρ_1 et ρ_2 , où ρ_1 et ρ_2 sont les deux racines (normalement distinctes) de $\rho^2 - 2x\rho + 1 = 0$. On trouve bien $\rho = x \pm \sqrt{x^2 - 1}$. On fixe la combinaison linéaire appropriée de ρ_1^n et ρ_2^n à partir des conditions initiales $T_0(x) = 1$ et $T_1(x) = x$. \square

Voir aussi *Hey, here's a nicer write up of the non-recurrence relation:*

<http://www2.edc.org/makingmath/mathprojects/patternsPoly/links/Warshauer.pdf>

<http://www.maths.lancs.ac.uk/gilbert/m245/node7.html>

Obtention directe de (15), extensions de Galois.

On peut reprendre (9) sous la forme: trouver y_n de degré n et $x_n \in \mathcal{P}_{n-1}$ tels que

$$y_n^2 - px_n^2 = \text{constante}, \quad (17)$$

où p est le polynôme $p(x) = x^2 - 1$.

(9) implique bien $y_n = T_n$ et $x_n = T'_n/n$ dans (17), mais on peut retrouver toutes les propriétés de T_n sans utiliser le fait que x_n est lié à la dérivée de y_n . Comme Tchebycheff lui-même¹⁰ [Tche], on considère (ce que l'on appelle maintenant) l'extension de Galois $\mathcal{G} = \{r + s\sqrt{p}, r, s \in \mathbb{C}\{x\}\}$ du corps des fonctions rationnelles. \mathcal{G} est bien un champ (= corps commutatif) : l'addition est triviale; multiplication: si r_1, r_2, s_1 et s_2 sont des fonctions rationnelles, $(r_1 + s_1\sqrt{p})(r_2 + s_2\sqrt{p}) = r_1r_2 + s_1s_2p + (r_1s_2 + r_2s_1)\sqrt{p}$; inversion: $\frac{1}{r + s\sqrt{p}} = \frac{r - s\sqrt{p}}{r^2 - s^2p}$. Un **entier** de \mathcal{G} est $r + s\sqrt{p}$ où r et s sont des polynômes. Une unité de \mathcal{G} est un entier dont l'inverse est encore un entier, donc, tel que $r^2 - s^2p = \text{constante}$, nous y voilà.

Avec $p(x) = x^2 - 1$, $x + \sqrt{x^2 - 1}$ et $x - \sqrt{x^2 - 1}$ sont des unités, puisqu'ils sont inverses l'un de l'autre. $(x \pm \sqrt{x^2 - 1})^n$ sont aussi des unités, donc de la forme $Y_n(x) \pm Z_n(x)\sqrt{x^2 - 1}$ où Y_n et Z_n sont des polynômes. A une constante multiplicative près, il n'y a pas d'autre unité $y_n(x) + x_n(x)\sqrt{x^2 - 1}$ avec y_n de degré n : de $y_n^2 - px_n^2 = \text{constante}$, on tire qu'une détermination de la racine carrée de $p(x) = x^2 - 1$ assure $y_n(x) - x_n(x)\sqrt{x^2 - 1} \sim \text{const.}/x^n$ quand $x \rightarrow \infty$. Soit $z = x + \sqrt{x^2 - 1}$ avec $z \rightarrow \infty$ quand $x \rightarrow \infty$. De $x = (z + z^{-1})/2$, $y_n(x) - x_n(x)\sqrt{x^2 - 1} = y_n\left(\frac{z + z^{-1}}{2}\right) - \frac{z - z^{-1}}{2}x_n\left(\frac{z + z^{-1}}{2}\right)$ se présente comme une combinaison de puissances $z^{-n}, z^{-n+1}, \dots, z^{n-1}, z^n$. Mais le comportement quand $x \rightarrow \infty$, donc quand $z \rightarrow \infty$ impose que seul le monôme en z^{-n} est présent.

On déduit alors (15) de $y_n(x) \pm x_n(x)\sqrt{x^2 - 1} = T_n(x) \pm U_{n-1}(x)\sqrt{x^2 - 1} = (x \pm \sqrt{x^2 - 1})^n$.

Quand p est de degré > 2 , on ne trouve plus d'unité $y_n \pm x_n\sqrt{p}$ non triviale, mais on a encore des polynômes tels que $y_n^2 - px_n^2$ soit de degré $\leq g$, où g (*genre*) est le plus petit entier tel que $2g + 2 \geq \text{degré de } p$ (théorie de Abel et Jacobi). Les fonctions rationnelles y_n/x_n sont alors de très intéressantes **approximations rationnelles** de \sqrt{p} .

¹⁰Grand merci à M. J. Meinguet d'avoir fourni cette information.

(4) **Fonction génératrice.**

$$\sum_{n=0}^{\infty} t^n T_n(x) = \frac{1-tx}{1-2tx+t^2}, \quad |t| < 1, -1 \leq x \leq 1. \quad (18)$$

Comme $-1 \leq x \leq 1$, $|T_n(x)| \leq 1$, la série du membre de gauche est absolument convergente. Multiplions-là par $1-2tx+t^2$ et réarrangeons:

$$\begin{aligned} (1-2tx+t^2) \sum_{n=0}^{\infty} t^n T_n(x) &= \sum_{n=0}^{\infty} t^n T_n(x) - 2tx \sum_{n=0}^{\infty} t^n T_n(x) + t^2 \sum_{n=0}^{\infty} t^n T_n(x) \\ &= \sum_{n=0}^{\infty} t^n T_n(x) - 2x \sum_{n=0}^{\infty} t^{n+1} T_n(x) + \sum_{n=0}^{\infty} t^{n+2} T_n(x) \\ &= 1-tx + \sum_{n=2}^{\infty} t^n [T_n(x) - 2xT_{n-1}(x) + T_{n-2}(x)] \\ &= 1-tx. \quad \square \end{aligned}$$

Voir aussi division de polynômes selon les puissances croissantes:
dividende = diviseur \times quotient + reste, pourvu que **l'ordre** (plus **petite** puissance figurant dans le polynôme) du reste soit **strictement supérieur à l'ordre** du diviseur.

Ici, on a, selon les puissances croissantes de t :

$$\begin{array}{rcl} 1-xt & = & (1-2xt+t^2) \quad 1 \quad + \quad xt-t^2 \\ xt-t^2 & = & (1-2xt+t^2) \quad xt \quad + \quad (2x^2-1)t^2-xt^3 \\ (2x^2-1)t^2-xt^3 & = & (1-2xt+t^2) \quad (2x^2-1)t^2 \quad + \quad (4x^3-3x)t^3-(2x^2-1)t^4 \end{array}$$

$$\text{etc.: } 1-xt = (1-2xt+t^2)[1+tT_1(x) + \dots + t^n T_n(x)] + t^{n+1} T_{n+1}(x) - t^{n+2} T_n(x).$$

Pour t fixé, la série (18) converge dans *l'ellipse* $|x + \sqrt{x^2-1}| < 1/|t|$: prendre θ complexe, la série est une combinaison des deux séries *géométriques* $\sum_{k \geq 0} t^k \exp(ik\theta)$ et $\sum_{k \geq 0} t^k \exp(-ik\theta)$, qui convergent donc tant que $|e^{i\theta}|$ et $|e^{-i\theta}| < 1/|t|$, donc tant que $|\text{Im } \theta| < \ln(1/|t|)$.

Cas particulier.¹¹ $\sum_{n=0}^{\infty} x^n T_n(x) \equiv 1, \quad -1 < x < 1.$

Que se passe-t-il si on ne garde qu'un nombre fini de termes (plus de problème de convergence!)? On a

$$(1-2tx+t^2) \sum_{n=0}^N t^n T_n(x) = 1-tx - t^{N+1} T_{N+1}(x) + t^{N+2} T_N(x).$$

Donc valable même si $|t| = 1$. Prenons $t = \exp(i\varphi)$, divisons par t^{N+2} , et soustrayons de la même expression en $1/t$:

$$(t^{-1}-2x+t) \sum_{n=0}^N (t^{N+1-n} - t^{n-N-1}) T_n(x) = t^{N+2} - t^{-2-N} - x(t^{N+1} - t^{-1-N}) + (t-t^{-1}) T_{N+1}(x),$$

enfin, avec $y = \cos \varphi$,

$$U_N(y) + 2 \sum_{n=1}^N U_{N-n}(y) T_n(x) = \frac{T_{N+1}(x) - T_{N+1}(y)}{x-y}. \quad (19)$$

¹¹Trouvé par S. Ambrose, DEA FSA, printemps 1999.

(5) **Formule de Christoffel-Darboux.**

$$\frac{T_{n+1}(x)T_n(y) - T_n(x)T_{n+1}(y)}{x - y} = 2 \sum_{k=0}^n T_k(x)T_k(y). \quad (20)$$

En effet, 1. Vrai si $n = 0$; 2. $n - 1 \rightarrow n$: on applique la relation de récurrence (12) pour se ramener à l'ancien numérateur

$$\begin{aligned} T_{n+1}(x)T_n(y) - T_n(x)T_{n+1}(y) &= [2xT_n(x) - T_{n-1}(x)]T_n(y) - T_n(x)[2yT_n(y) - T_{n-1}(y)] \\ &= 2(x - y)T_n(x)T_n(y) + T_n(x)T_{n-1}(y) - T_{n-1}(x)T_n(y) \end{aligned}$$

et on divise par $x - y$. □

(6) **Solutions d'équations du 3^{ème} degré par polynômes de Tchebycheff.** On porte

$x = -\frac{b}{3a} + t \frac{2\sqrt{b^2 - 3ac}}{3a}$ dans $ax^3 + bx^2 + cx + d = 0$ pour obtenir $4t^3 - 3t = T_3(t) = A$, où $A = [-2b^3 + 9abc - 27a^2d]/[2(b^2 - 3ac)^{3/2}]$. Les racines sont donc $t = \cos([\arccos A + 2k\pi]/3)$, $k = 0, 1, 2$. Formule algébrique: $t =$ les 3 déterminations de $T_{1/3}(A)$, donc, d'après (15), $t = (u + 1/u)/2$, où u est une des trois racines cubiques complexes de $A + \sqrt{A^2 - 1}$.

Voir "Chebyshev radicals" dans <http://www.answers.com/topic/cubic-equation>,
http://www-history.mcs.st-and.ac.uk/history/HistTopics/Quadratic_etc_equations.html

3.13. Equation différentielle linéaire.

Théorème. $y = T_n(x)$ est une solution de

$$(1 - x^2)y'' - xy' + n^2y = 0. \quad (21)$$

En effet, on peut dériver (9):

$$-2xy'^2 + 2(1 - x^2)y'y'' = -2n^2yy'$$

et diviser par y' (aux zéros de y' , (21) est établie par continuité de y , y' et y'').

$$\text{On peut aussi partir de (11) : } y = \cos n\theta \Rightarrow \frac{d^2y}{d\theta^2} = -n^2y, \quad (22)$$

et utiliser $x = \cos \theta$: $d/d\theta = d/d \arccos x = -\sqrt{1 - x^2} d/dx$,

$$\frac{d^2y}{d\theta^2} = \frac{d}{d\theta} \left(\frac{d}{d\theta} y \right) = \sqrt{1 - x^2} \left(\sqrt{1 - x^2} y' \right)',$$

d'où une forme intéressante (*formellement autoadjointe*)

$$\sqrt{1 - x^2} (\sqrt{1 - x^2} y')' = -n^2y. \quad (23)$$

équivalente à (21). □

3.14. Proposition. La solution générale de la récurrence (12) aussi bien que de l'équation différentielle (21) est

$$AT_n(x) + B\sqrt{1 - x^2}U_{n-1}(x) = A \cos n\theta + B \sin n\theta, \quad (24)$$

où U_{n-1} est le polynôme de Tchebycheff de deuxième espèce de degré $n - 1$, défini par

$$U_{n-1}(x) = \frac{\sin n\theta}{\sin \theta}, \quad (x = \cos \theta). \quad (25)$$

Lorsque (24) désigne la solution générale de la récurrence (12), A et B peuvent dépendre de x , mais pas de n ; si (24) désigne la solution générale de l'équation différentielle (21), A et B peuvent dépendre de n , mais pas de x .

(1) Récurrence:

En effet, on peut reprendre la démonstration de (12) et constater que toute expression $C \cos(n\theta + \varphi)$ convient, où C et φ peuvent dépendre de θ , donc de $x = \cos \theta$, mais pas de n . On retrouve ainsi $\cos n\theta$ avec $\varphi = 0$, mais aussi $\sin n\theta$ avec $\varphi = -\pi/2$, ou encore $\sin(n-1)\theta$ avec $\varphi = -(\theta + \pi/2)$. On obtient donc bien la dépendance annoncée en n .

On peut aussi reprendre la formule en puissance $n^{\text{èmes}}$ de ρ_1 et $\rho_2 = x \pm \sqrt{x^2 - 1} = \exp(\pm i\theta)$ (cf. p. 45) pour retrouver une combinaison de $\cos n\theta = T_n(x)$ et $\sin n\theta = \sqrt{1-x^2}U_{n-1}(x)$ avec des coefficients pouvant dépendre de x mais pas de n .

(2) Equation différentielle:

On voit bien que deux solutions indépendantes de (21) mise sous la forme (22) sont $\cos n\theta = T_n(x)$ et $\sin n\theta = \sqrt{1-x^2}U_{n-1}(x)$. \square

3.15. Polynômes de Tchebycheff et problèmes de Sturm-Liouville. L'équation (23) est bien de la forme de *Sturm-Liouville*

$$(p(x)y'(x))' + q(x)y(x) = \lambda w(x)y(x), \quad a < x < b, \quad (26)$$

avec $p(x) > 0$ et $w(x) > 0$ sur (a, b) . Un *problème de Sturm-Liouville régulier* consiste à trouver des solutions non nulles (*fonctions propres*) de (26) vérifiant des conditions aux limites homogènes, si $p(a)$ et $p(b) \neq 0$. Ici, $p(x) = \sqrt{1-x^2}$, on a un *problème singulier*: $p(a) = p(b) = 0$. Les fonctions propres sont alors simplement définies par des conditions de régularité aux limites: valeurs et dérivées finies quand $x \rightarrow a$ et b .

Ici, avec $p(x) = \sqrt{1-x^2}$, $q(x) = 0$ et $w(x) = 1/\sqrt{1-x^2}$, la solution générale de (26) est une combinaison de $\cos \mu\theta$ et $\sin \mu\theta$, où $\lambda = -\mu^2$ (comme d'habitude, faire $x = \cos \theta$ pour obtenir $d^2y/d\theta^2 = -\mu^2y$). Les dérivées en x sont les dérivées en θ divisées par $-\sin \theta = -\sqrt{1-x^2}$, et restent donc finies en ± 1 si les dérivées en θ sont nulles en $\theta = 0$ et π . Ceci impose $\sin \mu\pi = 0 \Rightarrow \mu \in \mathbb{Z}$, donc, $\lambda = -n^2$, $n = 0, 1, \dots$ et la fonction propre correspondante = $\cos n\theta = T_n(x)$.

Pour traiter convenablement les problèmes de Sturm-Liouville, il faut les définir sur des *espaces de Hilbert* (*espaces de Sobolev*).

Dérivées d'ordre quelconque: en dérivant (21) p fois, on obtient

$$(1-x^2)\frac{d^{p+2}T_n(x)}{dx^{p+2}} - (2p+1)x\frac{d^{p+1}T_n(x)}{dx^{p+1}} + (n^2-p^2)\frac{d^pT_n(x)}{dx^p} = 0, \quad (27)$$

(par récurrence [sur p], ou en utilisant une formule de Leibniz de la dérivée $p^{\text{ème}}$ d'un produit), ou encore

$$\frac{d}{dx} \left((1-x^2)^{p+1/2} \frac{d^{p+1}T_n(x)}{dx^{p+1}} \right) + (n^2-p^2)(1-x^2)^{p-1/2} \frac{d^pT_n(x)}{dx^p} = 0. \quad (28)$$

3.16. Coefficients, dérivées et primitives.

(1) $T_n(x) = \sum_{\substack{j \geq 0 \\ 2j \leq n}} \frac{(-1)^j n(n-j-1)! 2^{n-1-2j}}{j!(n-2j)!} x^{n-2j}$. En effet, nous savons déjà que le coefficient de x^n est 2^{n-1} , et que les coefficients non nuls sont ceux de même parité que n . Ensuite, par (27) avec $p = n - 2j$, le coefficient de x^{n-2j} est $1/(n-2j)!$ fois

$$\begin{aligned} \frac{d^{n-2j} T_n(0)}{dx^{n-2j}} &= \frac{-1}{n^2 - (n-2j)^2} \frac{d^{n-2j+2} T_n(0)}{dx^{n-2j+2}} \\ &= \frac{-1}{4j(n-j)} \frac{d^{n-2j+2} T_n(0)}{dx^{n-2j+2}} \\ &= \frac{(-1)^j}{4^j j(j-1) \cdots 1 (n-j)(n-j+1) \cdots (n-1)} 2^{n-1} n! \quad \square \end{aligned}$$

Le rapport des valeurs absolues des coefficients de x^{n-2j} et de x^{n-2j+2} est

$$\frac{(n-2j+2)(n-2j+1)}{4j(n-j)} \approx \frac{1}{4\alpha(\alpha+1)} \text{ où } \alpha = j/(n-2j). \text{ Ce rapport est donc } > 1$$

tant que $\alpha < (\sqrt{2}-1)/2$. A ce moment, c'est-à-dire quand $j \sim n(2-\sqrt{2})/4$, on est en présence du plus grand coefficient, qui est de l'ordre de $(\sqrt{2}+1)^n$ (par Stirling, voir p. 182)

(2) Plus intéressante est l'expression des puissances de x dans la base des polynômes de Tchebycheff, base que l'on va s'habituer à utiliser dans la suite:

$$x^n = \frac{1}{2^{n-1}} \left(\sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{k} \frac{T_{n-2k}(x)}{1 + \delta_{n-2k,0}} \right),$$

($\lfloor \cdot \rfloor$ = partie entière par défaut),

où $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ est le nombre de combinaisons de n objets pris k à k (coefficient binomial).

En effet, $x = \cos \theta = (e^{i\theta} + e^{-i\theta})/2$, $x^n = 2^{-n} \sum_{k=0}^{k=n} \binom{n}{k} e^{i(n-2k)\theta}$,

(3)

$$T'_n = nU_{n-1} = 2n \left(T_{n-1} + T_{n-3} + \cdots + T_1 \text{ ou } \frac{T_0}{2} \right). \quad (29)$$

$$\begin{aligned} T'_n(x) &= \frac{dT_n(x)}{dx} = \frac{d \cos n\theta}{d \cos \theta} = \\ &= n \frac{\sin n\theta}{\sin \theta} = nU_{n-1}(x) \\ &= n \frac{e^{in\theta} - e^{-in\theta}}{e^{i\theta} - e^{-i\theta}} \\ &= n(e^{i(n-1)\theta} + e^{i(n-3)\theta} + \cdots + e^{-i(n-3)\theta} + e^{-i(n-1)\theta}) \\ &= 2n(T_{n-1}(x) + T_{n-3}(x) + \cdots) \end{aligned}$$

où on termine sur $e^{i\theta} + e^{-i\theta} = 2T_1(x)$ si n est pair, et sur $e^{i0\theta} = 1 = T_0$ si n est impair. \square

Remarquons que l'on a aussi montré $U_n(x) - U_{n-2}(x) = 2T_n(x)$.

(4)

$$\int T_n(x)dx = \frac{T_{n+1}(x)}{2(n+1)} - \underbrace{\frac{T_{n-1}(x)}{2(n-1)}}_{\text{si } n > 1} + C. \tag{30}$$

$$\begin{aligned} \int T_n(x)dx &= - \int \cos n\theta \sin \theta d\theta \\ &= -\frac{1}{2} \int [\sin(n+1)\theta - \sin(n-1)\theta]d\theta \\ &= \frac{\cos(n+1)\theta}{2(n+1)} - \frac{\cos(n-1)\theta}{2(n-1)} \quad (\text{si } n \neq 1) + C \end{aligned}$$

(5)

$$\int \frac{T_n(x)}{\sqrt{1-x^2}} dx = \frac{T_{n+1}(x) - T_{n-1}(x)}{2n\sqrt{1-x^2}} + C \tag{31}$$

$$\begin{aligned} \int \frac{T_n(x)}{\sqrt{1-x^2}} dx &= - \int \cos n\theta d\theta = -\frac{\sin n\theta}{n} + C \\ &= -\frac{\sin n\theta \sin \theta}{n \sin \theta} + C = \frac{\cos(n+1)\theta - \cos(n-1)\theta}{2n \sin \theta} + C \quad (n > 0) \end{aligned}$$

(6) **Exercice: primitive de $T_n(x) \log(x-c)$.**

La primitive est de la forme $A_n(x) \log(x-c) - B_n(x)$, où A_n est une primitive de T_n et B_n est un polynôme. Choisissons pour A_n la primitive de T_n qui s'annule en $x=c$, soit, par (30), $A_n(x) = \frac{T_{n+1}(x) - T_{n+1}(c)}{2(n+1)} - \frac{T_{n-1}(x) - T_{n-1}(c)}{2(n-1)}$. $B_n(x)$ est alors une primitive de $A_n(x)/(x-c)$.

D'après (19), $\frac{A_n(x)}{x-c} = \frac{U_n(c)}{2(n+1)} + \sum_{k=1}^n \frac{U_{n-k}(c)}{n+1} T_k(x) - \frac{U_{n-2}(c)}{2(n-1)} - \sum_{k=1}^{n-2} \frac{U_{n-2-k}(c)}{n-1} T_k(x)$.

Donc, $B_n(x) = x \frac{U_n(c)}{2(n+1)} + \sum_{k=1}^n \frac{U_{n-k}(c)}{n+1} A_k(x) - x \frac{U_{n-2}(c)}{2(n-1)} - \sum_{k=1}^{n-2} \frac{U_{n-2-k}(c)}{n-1} A_k(x)$.

3.17. Primitives itérées de $\frac{T_n(x)}{\sqrt{1-x^2}}$ et formule de Rodrigues.

Reprenons (31) et appelons $V_{n,1}(x) := \int_{-1}^x V_{n,0}(t) dt := \int_{-1}^x \frac{T_n(t)}{\sqrt{1-t^2}} dt \quad (n > 0)$. On a donc vu que $V_{n,1}$

est une combinaison des $V_{.,0}$: $V_{n,1} = \frac{V_{n+1,0}}{2n} - \frac{V_{n-1,0}}{2n}$, et que $V_{n,1}(x) = -\sin(n\theta)/n$. $V_{n,1}$ est une fonction continue, bornée par $1/n$ sur $[-1, 1]$, nulle en ± 1 .

Les intégrales successives

$$V_{n,p}(x) := \int_{-1}^x V_{n,p-1}(t) dt, \quad p = 1, 2, \dots, n \geq p \tag{32}$$

sont donc encore des combinaisons des $V_{.,0}$, très exactement de $V_{n-p,0}, \dots, V_{n+p,0}$, ou encore, si $p \geq 1$,

de $V_{n-p+1,1}, \dots, V_{n+p-1,1}$, soit $V_{n,p}(x) = \sum_{k=n-p+1}^{n+p-1} \alpha_{k,n}^{(p)} \sin(k\theta)$, On s'intéresse particulièrement aux sommes

$S_n^{(p)} = \sum_{k=n-p+1}^{n+p-1} |\alpha_{k,n}^{(p)}|$, bornes supérieures de $\|V_{n,p}\|_\infty$. On trouve **Proposition.** *Les intégrales successives*

(32) *admettent les bornes*

$$|V_{n,p}(x)| \leq \frac{1}{(n-p+1)(n-p+3) \cdots (n+p-3)(n+p-1)}, \quad -1 \leq x \leq 1, \quad p = 1, 2, \dots, n \geq p. \tag{33}$$

On a aussi la forme intéressante

Proposition. *Les intégrales successives (32) sont données par*

$$V_{n,p}(x) = \frac{(-1)^p}{n^2(n^2-1)\cdots(n^2-(p-1)^2)}(1-x^2)^{p-1/2} \frac{d^p T_n(x)}{dx^p}, \quad p = 0, 1, \dots, n. \quad (34)$$

En effet, on passe de p à $p+1$ par (28). En $p=n$,

$$(1-x^2)^{-1/2} T_n(x) = \frac{(-1)^n 2^n n!}{(2n)!} \frac{d^n}{dx^n} \left((1-x^2)^{n-1/2} \right), \quad (35)$$

une formule de Rodrigues¹² (cf. p. 104, p. 105).

3.18. Orthogonalité et base duale de \mathcal{P}_N^* .

3.18.1. Orthogonalité des polynômes de Tchebycheff.

$$\int_{-1}^1 T_n(x) T_m(x) \frac{dx}{\sqrt{1-x^2}} = \begin{cases} \pi & \text{si } m = n = 0, \\ \frac{\pi}{2} & \text{si } m = n \neq 0 \\ 0 & \text{si } m \neq n \end{cases} \quad (36)$$

La vérification est immédiate, en passant comme d'habitude par les fonctions trigonométriques:

nous devons évaluer $\int_0^\pi \cos(m\theta) \cos(n\theta) d\theta = \frac{1}{2} \int_0^\pi [\cos((m+n)\theta) + \cos((m-n)\theta)] d\theta$, et

on utilise $\int_0^\pi \cos(p\theta) d\theta = \pi$ si $p=0$; $= \frac{\sin(p\pi) - \sin 0}{p} = 0$ si p est un entier non nul. \square

On peut aussi utiliser l'

3.18.2. *orthogonalité de deux fonctions propres relatives à deux valeurs propres distinctes du problème de Sturm-Liouville.* (26): si y_n est une solution de

$$(p(x)y_n'(x))' + q(x)y_n(x) = \lambda_n w(x)y_n(x), \quad a < x < b,$$

avec $p(x)y_n(x) = 0$ en $x = a$ et $x = b$, multiplions par $y_m(x)$ et intégrons par parties:

$$[p(x)y_n'(x)y_m(x)]_a^b - \int_a^b p(x)y_n'(x)y_m'(x) dx + \int_a^b q(x)y_n(x)y_m(x) dx = \lambda_n \int_a^b w(x)y_n(x)y_m(x) dx,$$

en remarquant que la contribution des termes aux limites est nulle. Reprenons en intervertissant m et n , et soustrayons. Il vient $\int_a^b w(x)y_n(x)y_m(x) dx = 0$ si $m \neq n$.

3.18.3. *Base duale de \mathcal{P}_N^* .* Toute forme λ sur un espace vectoriel V de dimension finie est entièrement décrite par son action sur tous les éléments d'une base $\{b_0, b_1, \dots, b_N\}$ de V . En effet, $\lambda(f)$ pour $f \in V$ s'obtient à partir de la représentation de f dans la base choisie:

$$f = \sum_0^N c_k b_k \Rightarrow \lambda(f) = \sum_0^N c_k \lambda(b_k).$$

¹²La première parution de (35) est au bas de la p. 4 de C.G.J. Jacobi, Formula transformationis integralium definitorum, *J. reine angew. Math.* **15** (1836) 1–38 (remarque aimablement communiquée par A. Ronveaux, nov. 2003).

Des formes $\lambda_0, \lambda_1, \dots$ sont donc entièrement décrites par les vecteurs $[\lambda_0(b_0), \dots, \lambda_0(b_N)], [\lambda_1(b_0), \dots, \lambda_1(b_N)], \dots$ de \mathbb{R}^{N+1} ou \mathbb{C}^{N+1} . Le dual V^* de V est donc également un espace vectoriel de même dimension que celle de V .

Les coefficients c_k sont eux-mêmes les résultats de l'action de formes de V^* sur f ! Ces formes c_0, \dots, c_N constituent ce que l'on appelle la **base duale** de V^* relativement à la base $\{b_0, \dots, b_N\}$ de V . On a $c_m(b_n) = \delta_{m,n}$, on dit que $\{c_m\}_0^N$ est la base de V^* **biorthogonale** à la base $\{b_n\}_0^N$ de V .

La propriété d'orthogonalité (36) des polynômes de Tchebycheff permet d'explicitier la base de \mathcal{P}_N^* duale relativement à la base des polynômes de Tchebycheff de \mathcal{P}_N :

Proposition. La base de \mathcal{P}_N^* duale relativement à la base $\left\{ \frac{T_0}{2}, T_1, \dots, T_N \right\}$ de \mathcal{P}_N est

$$c_k : \quad c_k(f) = \frac{2}{\pi} \int_{-1}^1 f(x) T_k(x) \frac{dx}{\sqrt{1-x^2}} = \frac{2}{\pi} \int_0^\pi f(\cos \theta) \cos(k\theta) d\theta, \quad (37)$$

pour $k = 0, \dots, N$.

En effet, vérifions que $c_m(b_n) = \delta_{m,n}$: 1.) avec $b_0 = T_0/2$, on a $c_m(b_0) = \frac{2}{\pi} \int_{-1}^1 \frac{T_0}{2} T_m(x) \frac{dx}{\sqrt{1-x^2}} = \delta_{m,0}$, par (36) avec $n = 0$; 2.) avec $b_n = T_n$, $n > 0$, $c_m(b_n) = \frac{2}{\pi} \int_{-1}^1 T_n(x) T_m(x) \frac{dx}{\sqrt{1-x^2}} = \delta_{m,n}$, par (36) avec $n > 0$. \square

L'avantage d'avoir pris $T_0/2$ au lieu de T_0 dans la base de \mathcal{P}_N est d'avoir une formule uniforme dans (37). On a donc

$$\forall f \in \mathcal{P}_N, \quad f = \frac{c_0(f)}{2} T_0 + c_1(f) T_1 + \dots + c_N(f) T_N := \sum'_{k=0}^N c_k(f) T_k, \quad (38)$$

avec $c_k(f)$ donné par (37); la notation \sum' signifiant que la somme s'effectue en prenant la moitié du premier terme indiqué, les autres étant inchangés.

On a d'ailleurs encore un autre éclairage très intéressant de (38) comme **somme de Fourier**: $\forall f \in \mathcal{P}_N$,

$$f(\cos \theta) = \frac{c_0(f)}{2} + c_1(f) \cos \theta + \dots + c_N(f) \cos(N\theta) = \sum'_{j=0}^N c_j(f) \cos(j\theta) = \frac{1}{2} \sum_{j=-N}^N c_{|j|} e^{ij\theta} \quad (39)$$

3.18.4. *Orthogonalité discrète des T_n .* . Comme on a vu en p. 21 que toute forme sur un espace de dimension finie peut s'exprimer comme $\varphi(f) = \sum_0^N \alpha_m f(x_m)$, il doit en être de même pour (37). On dispose de deux formules particulièrement utiles:

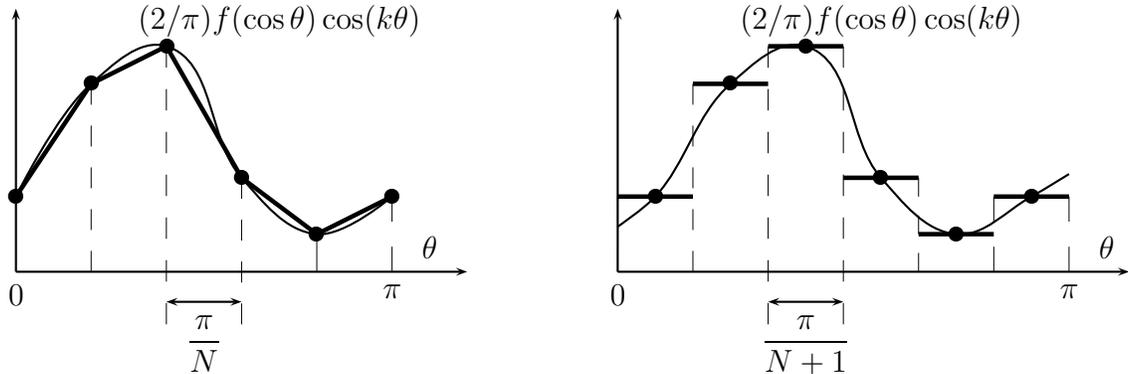
Proposition. Si $N > 0$, on a les deux formules discrètes valables pour $\forall f \in \mathcal{P}_N$

$$\text{(Trapèzes:)} \quad c_k(f) = \frac{2}{N} \sum''_{m=0}^N f\left(\cos \frac{m\pi}{N}\right) \cos\left(\frac{km\pi}{N}\right). \quad (40)$$

\sum'' désignant une somme où le premier et le dernier terme sont divisés par 2;

$$\text{(Rectangles:)} \quad c_k(f) = \frac{2}{N+1} \sum_{m=0}^N f\left(\cos \frac{(m+1/2)\pi}{N+1}\right) \cos\left(\frac{k(m+1/2)\pi}{N+1}\right). \quad (41)$$

Les dénominations 'trapèzes' et 'rectangles' viennent de l'interprétation de (40) et (41) comme somme d'aires de trapèzes ou de rectangles équivalentes à l'intégrale (37) donnant l'aire sous le graphe de la fonction $2f(\cos \theta) \cos(k\theta)/\pi$:



Ces formules d'intégration, apparemment très rudimentaires, sont donc **exactes** pour $f \in \mathcal{P}_N$!

Démonstration de (40) et (41). Entrons (39) dans (40) et (41): nous devons vérifier

$$c_k(f) = \sum_{m=m_1}^{m_2} \mu_m \left[\frac{1}{2} \sum_{j=-N}^N c_{|j|} e^{ij\theta_m} \right] \cos(k\theta_m),$$

où les μ_m et les θ_m dépendent de la formule discrète, il nous suffit de savoir que les θ_m sont en progression arithmétique. Par parité du cosinus, on peut d'ailleurs considérer que l'on a $\mu_m = 1/N, m_1 = -N, m_2 = N - 1$ dans (40); $\mu_m = 1/(N + 1), m_1 = -N - 1, m_2 = N$ dans (41). La double somme se réarrange en

$$\frac{1}{2} \sum_{j=-N}^N c_{|j|} \left[\sum_{m=m_1}^{m_2} \mu_m e^{ij\theta_m} \cos(k\theta_m) \right] = \frac{1}{4} \sum_{j=-N}^N c_{|j|} \left[\sum_{m=m_1}^{m_2} \mu_m (e^{i(j+k)\theta_m} + e^{i(j-k)\theta_m}) \right]$$

La somme intérieure est une progression géométrique de raison $\exp(i(j \pm k)\Delta)$, où Δ est la raison de la progression arithmétique des θ_m (π/N ou $\pi/(N + 1)$). Cette raison vaut l'unité quand $j = \mp k$, la somme intérieure vaut alors μ_m fois le nombre de termes = 2; si $j \neq \mp k$, la progression géométrique vaut une expression de numérateur raison^{1+nombre de termes} - 1, ce qui vaut 0 car la raison élevée à une puissance égale au nombre de termes +1 donne $\exp(2\pi i) = 1$. \square

Si on prend $f = T_n$, on a d'ailleurs des formules d'orthogonalité discrète dont voici le détail:

$$\sum_{m=0}^N T_p \left(\cos \frac{m\pi}{N} \right) T_q \left(\cos \frac{m\pi}{N} \right) = \begin{cases} N & \text{si } p + q \text{ et } p - q \text{ sont} \\ & \text{des multiples de } 2N, \\ \frac{N}{2} & \text{si un seul des deux nombres } p + q \text{ ou } p - q \\ & \text{est multiples de } 2N \\ 0 & \text{dans les autres cas} \end{cases} \quad (42)$$

$$\sum_{m=0}^{N-1} T_p \left(\cos \frac{(m+1/2)\pi}{N} \right) T_q \left(\frac{(m+1/2)\pi}{N} \right) = \begin{cases} N & \text{si } p+q \text{ et } p-q \text{ sont des} \\ & \text{multiples de } 4N, \\ -N & \text{si } p+q \text{ et } p-q \text{ sont des} \\ & \text{multiples impairs de } 2N, \\ (-1)^{(p\pm q)/(2N)} \frac{N}{2} & \text{si un seul des deux nombres } p+q \\ & \text{ou } p-q \text{ est multiple de } 2N \\ 0 & \text{dans les autres cas} \end{cases} \quad (43)$$

4. Bonne approximation; séries de polynômes de Tchebycheff, relation avec séries de Fourier.

On a vu apparaître les polynômes de Tchebycheff en résolvant le problème très particulier de la meilleure approximation d'un polynôme par un polynôme de degré immédiatement inférieur.

On va décrire maintenant une manière d'utiliser cette technique particulière pour construire une approximation polynomiale raisonnable (*bonne approximation*) d'une fonction continue arbitraire.

4.1. Bonne base et bonne approximation.

Soit $\{b_1, \dots, b_N\}$ une base d'un espace V_N de dimension $N < \infty$, de sorte que $\forall f \in V_N$,

$$f = \sum_{k=1}^N c_k(f) b_k.$$

La base est bonne relativement à une norme $\| \cdot \|$ si

$$C_N \leq \frac{\sum_1^N |c_k(f)|}{\|f\|} \leq D_N,$$

avec D_N/C_N pas trop grand, par exemple, une puissance faiblement croissante de N .

Remarquons que $f = b_n \Rightarrow 1/D_n \leq \|b_n\| \leq 1/C_n$.

La base $\left\{ \frac{T_0}{2}, T_1, \dots, T_N \right\}$ est une bonne base de \mathcal{P}_N relativement à $\| \cdot \|_\infty$ puisque, d'après

$$(37), |c_k(f)| \leq \frac{2}{\pi} \int_0^\pi |f(\cos \theta)| |\cos k\theta| d\theta \leq 2\|f\|_\infty, \text{ donc}$$

$$|c_k(f)| \leq 2\|f\|_\infty,$$

$$1 \leq \frac{\sum_1^N |c_k(f)|}{\|f\|_\infty} \leq 2N + 1.$$

La base des puissances de x n'est **pas du tout** une bonne base relativement à la norme $\| \cdot \|_\infty$ sur $[-1, 1]$! On a bien $C = 1$, mais D vaut au moins 2^{N-1} (prendre $f = T_N$).

Par contre, la base des puissances de x est une excellente base sur le **disque unité** du plan complexe.

Soit V_n le sous-espace de V_N sous-tendu par b_1, \dots, b_n .

Si b_1, \dots, b_N est une bonne base de V_N , $p_n = \sum_1^n c_k(f)b_k$ est une bonne approximation de f dans V_n :

$$\|f - p_n\| = \left\| \sum_{n+1}^N c_k(f)b_k \right\| = \left\| \sum_{n+1}^N c_k(f - \hat{p}_n)b_k \right\| \leq C_N^{-1} \sum_{n+1}^N |c_k(f - \hat{p}_n)| \leq \frac{D_N}{C_N} \|f - \hat{p}_n\|.$$

(Cf. lemme de Lebesgue, p. 58).

4.2. Cascade de meilleures approximations et développements dans la base des polynômes de Tchebycheff.

Voici une autre façon de suggérer que la somme partielle de degré n du développement d'une fonction en polynômes de Tchebycheff est proche de la meilleure approximation de degré $\leq n$:

Partons d'un polynôme p de degré N . Nous savons construire la meilleure approximation de degré $\leq N-1$ de $p \in \mathcal{P}_N$: on l'obtient en soustrayant de p un multiple scalaire approprié de T_N de façon à aboutir à un degré $< N$. Appelons \mathcal{M}_N l'opérateur $\mathcal{P}_N \rightarrow \mathcal{P}_{N-1}$ qui réalise cette extraction de meilleure approximation: $\forall p \in \mathcal{P}_N$, $\mathcal{M}_N(p) = p - \frac{a_N(p)}{a_N(T_N)} T_N$, où $a_k(p)$

est, rappelons-le, le coefficient de x^k de p . Le multiplicateur de T_N utilisé dans la construction de $\mathcal{M}_N(p)$ est évidemment $c_N(p)$, le coefficient de T_N dans le développement de p dans la base des polynômes de Tchebycheff (38). Appliquons maintenant \mathcal{M}_{N-1} à $\mathcal{M}_N(p) \in \mathcal{P}_{N-1}$: $\mathcal{M}_{N-1}\mathcal{M}_N(p) = \mathcal{M}_{N-1}(p - c_N(p)T_N) = p - c_N(p)T_N - c_{N-1}(p)T_{N-1}$.

$\mathcal{M}_{N-1}\mathcal{M}_N(p)$ n'est normalement *pas* la meilleure approximation de p dans \mathcal{P}_{N-2} , mais comme $\mathcal{M}_{N-1}\mathcal{M}_N(p)$ diffère de p d'une combinaison de deux polynômes de Tchebycheff de degrés élevés, on peut penser que la fonction d'erreur n'est pas trop éloignée d'une fonction équi-oscillante.

En considérant des déflations successives $\mathcal{M}_k\mathcal{M}_{k+1} \cdots \mathcal{M}_N(p)$, on voit apparaître le développement de p dans la base des polynômes de Tchebycheff:

$$\mathcal{M}_{n+1}\mathcal{M}_{n+2} \cdots \mathcal{M}_N(p) = p - \sum_{k=n+1}^N c_k(p)T_k = \sum_{k=0}^n c_k(p)T_k \quad \text{si } p = \sum_{k=0}^N c_k(p)T_k.$$

Donc: la succession $\mathcal{S}_n(p) := \mathcal{M}_{n+1}\mathcal{M}_{n+2} \cdots \mathcal{M}_N(p)$ de meilleures approximations $((\cdots (\mathcal{P}_N \rightarrow \mathcal{P}_{N-1}) \cdots) \rightarrow \mathcal{P}_{n+1}) \rightarrow \mathcal{P}_n$ est ce qu'on appelle ici la **bonne approximation** $\mathcal{S}_n(p)$ de degré $\leq n$ de $p \in \mathcal{P}_N$. Elle consiste exactement en la somme des termes en T_0, T_1, \dots, T_n du développement de p dans la base $\{T_0, T_1, \dots, T_N\}$ de \mathcal{P}_N .

Il faudra encore définir ce que c'est une bonne approximation d'une fonction continue f , donc ce que devrait être $c_k(f)$.

Constatons en tout cas que (39) est une **somme partielle de Fourier** de la fonction périodique $F : F(\theta) := f(\cos \theta)$:

$$F(\theta) = \sum_{j=0}^N c_j(f) \cos(j\theta) = \frac{1}{2} \sum_{j=-N}^N c_{|j|}(f) e^{ij\theta}$$

4.3. Série de Fourier d'une fonction continue périodique.

Il existe une théorie très riche¹³ du développement en série de Fourier d'une fonction périodique sur \mathbb{R} . Soit F périodique de période 2π : $F(x+2\pi) = F(x)$, $\forall x \in \mathbb{R}$. Il est question d'établir quand la suite des **sommes partielles de Fourier**

$$\mathcal{S}_n(F)(\theta) := \frac{A_0(F)}{2} + \sum_{k=1}^n [A_k(F) \cos(k\theta) + B_k(F) \sin(k\theta)] = \frac{1}{2} \sum_{-n}^n C_k(F) e^{ik\theta}$$

converge vers F , où $C_k(F) := \pi^{-1} \int_{-\pi}^{\pi} F(\varphi) e^{-ik\varphi} d\varphi$, $k \in \mathbb{Z}$.

Avec cette définition, $C_k(F)$ coïncide avec le coefficient de Tchebycheff $c_k(f)$ de (37) si $F(\theta) = f(\cos \theta)$:

$$c_k(f) = \frac{2}{\pi} \int_0^{\pi} f(\cos \theta) \cos(k\theta) d\theta = \frac{1}{\pi} \int_{-\pi}^{\pi} F(\theta) \cos(k\theta) d\theta = \frac{1}{\pi} \int_{-\pi}^{\pi} F(\theta) e^{ik\theta} d\theta = C_{\pm k}(F)$$

car, F étant ici une fonction paire (en θ), $\int_{-\pi}^{\pi} F(\theta) \sin(k\theta) d\theta = 0$.

Nos sommes de Tchebycheff (38) et (39) sont donc bien des sommes de Fourier de F . Par l'argumentation développée dans le § 4.2, nous pouvons soupçonner que cette somme $\mathcal{S}_n(f)$ n'est pas éloignée de la meilleure approximation de f dans \mathcal{P}_n , ou encore que $\mathcal{S}_n(F)$ approche bien la fonction périodique F .

Pour plus de rigueur, et aussi afin de pouvoir quantifier l'erreur $f - \mathcal{S}_n(f) = F - \mathcal{S}_n(F)$, nous devons d'abord établir des **conditions de convergence de la série de Fourier** de F .

Le théorème le plus utile en cette matière est

Théorème de Dirichlet. *Les sommes partielles de Fourier d'une fonction continue périodique à variation bornée sur une période convergent uniformément vers la fonction.*

Voir R. Godement, *Analyse mathématique II*, Springer, 1998,
<http://www.pourlascience.com/numeros/pls-254/livres.htm>

Une fonction F est à variation bornée sur l'intervalle I si $V = \sup_{x_j \in I} \sum_j |F(x_{j+1}) - F(x_j)| < \infty$.

La fonction $F(\theta) = \theta \sin(1/\theta)$ est continue, mais de variation non bornée sur un intervalle contenant 0. Une fonction croissante bornée est à variation bornée; toute fonction réelle à variation bornée sur un intervalle est la différence de deux fonctions croissantes.

Le caractère à variation bornée est même ici plus important que le caractère continu, puisqu'une fonction périodique à variation bornée, éventuellement discontinue, est encore la limite ponctuelle de ses somme de Fourier (théorème de **Jordan**).

N.B. Une fonction \mathcal{C}^1 , ou même seulement Lipschitzienne, sur I est forcément à variation bornée: $V = \sup_{x_j \in I} \sum_j |F(x_{j+1}) - F(x_j)| \leq \text{const.} \times \text{longueur de } I$.

Sans démontrer complètement ici le théorème de Dirichlet, voyons avec R. Godement (pp. 282–293) quelques énoncés voisins:

Lemme de Riemann-Lebesgue. *Les coefficients de Fourier d'une fonction périodique intégrable tendent vers 0.*

¹³M. Willem, *Analyse harmonique réelle*, Hermann, Paris, 1995.

Sans démontrer ce lemme dans toute sa généralité, nous voyons déjà qu'il vaut pour la classe plus restreinte des fonctions de carré intégrable. Par l'inégalité de Bessel (chap. 3, § 7.1, p. 129), $\sum_{-\infty}^{\infty} |C_k(G)|^2 \leq \frac{2}{\pi} \int_{-\pi}^{\pi} |G(\theta)|^2 d\theta \Rightarrow C_k(G) \rightarrow 0, k \rightarrow \pm\infty$.

Détaillons maintenant la somme de Fourier de F :

$$\begin{aligned} \mathcal{S}_n(F)(\theta) &= \frac{1}{2} \sum_{-n}^n C_k(F) e^{ik\theta} \\ &= \frac{1}{2\pi} \sum_{-n}^n \int_{-\pi}^{\pi} F(\varphi) e^{-ik\varphi} d\varphi e^{ik\theta} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} F(\varphi) \left[\sum_{-n}^n e^{ik(\theta-\varphi)} \right] d\varphi \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} F(\varphi) \frac{e^{i(n+1)(\theta-\varphi)} - e^{-in(\theta-\varphi)}}{e^{i(\theta-\varphi)} - 1} d\varphi \\ \mathcal{S}_n(F)(\theta) - F(\theta) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{F(\theta + \varphi) - F(\theta)}{e^{-i\varphi} - 1} [e^{-i(n+1)\varphi} - e^{in\varphi}] d\varphi, \end{aligned}$$

où on a utilisé la périodicité de F et l'identité \mathcal{S}_n appliqué à une constante = cette même constante (ici, θ est fixé, donc $F(\theta)$ est constante).

On a donc $\mathcal{S}_n(F)(\theta) - F(\theta) = C_{n+1}(G) - C_{-n}(G)$, où G est la fonction qui vaut $G(\varphi) = \frac{1}{2} \frac{F(\theta + \varphi) - F(\theta)}{e^{-i\varphi} - 1}$ en φ (rappelons que θ est fixe). Le résultat tendra effectivement vers 0 quand $n \rightarrow \infty$ si F est Lipschitzienne.

Une série de Fourier divergente¹⁴ (du Bois-Reymond, Fejér): la fonction $F(\theta) = \sum_{m=1}^{\infty} \frac{\sin(2^{1+m^3}\theta)}{m^2} \sum_{p=1}^{2^{m^3}} \frac{\sin(p\theta)}{p}$

est périodique et continue (les sommes partielles de Fourier $\sum_{p=1}^N \frac{\sin(p\theta)}{p}$ de la fonction $\text{sign } \theta(\pi - |\theta|)/2$ étant toutes bornées par une même constante [le fameux phénomène de Gibbs]), mais son développement en série de Fourier est divergent en $\theta = 0$. En effet, les coefficients sont $C_{\pm k}(F) = \frac{1}{2m^2(2^{1+m^3} - k)}$ si $2^{m^3} \leq k < 2^{1+m^3}$, $-\frac{1}{2m^2(k - 2^{1+m^3})}$ si $2^{1+m^3} < k \leq 3 \cdot 2^{m^3}$, nuls pour les autres valeurs de k . La série des C_k diverge, puisque les sommes de C_k pour $k = 2^{m^3}$ à $k = 2^{1+m^3} - 1$ sont supérieures à $m/2$ (la somme $1 + 1/2 + \dots + 1/(2^N)$ est supérieure à $N/2$). Les sommes de C_k pour $k = 2^{1+m^3} + 1$ à $k = 3 \cdot 2^{m^3} - 1$ sont d'ailleurs tout aussi violemment négatives. Inutile d'ajouter que F est furieusement non dérivable en $\theta = 0 \dots$

4.4. Séries de polynômes de Tchebycheff.

Soit f continue sur $[-1, 1]$. On adoptera naturellement (38) avec $c_k(f)$ calculés par (37):

$$\mathcal{S}_n(f) = \sum_{k=0}^n c_k(f) T_k \quad (44)$$

¹⁴ D.C Champeney, *A Handbook of Fourier Theorems*, Cambridge U.P., 1987, § 5.5 et 15.2; R.E. Edwards, *Fourier Series, A Modern Introduction*, Holt Rinehart Winston 1967 (2^{ème} éd.: Springer, 1979), § 10.3.1.

comme définition de l'opérateur de bonne approximation appliqué à une fonction continue arbitraire f .

Ceci suggère que l'opérateur de bonne approximation découle d'un développement infini (série) applicable à toute fonction continue sur $[-1, 1]$. La définition suivante comporte pourtant une restriction:

Définition. *Le développement en série de polynômes de Tchebycheff d'une fonction continue sur $[-1, 1]$ est*

$$f = \sum_{k=0}^{\infty} c_k(f) T_k, \tag{45}$$

avec les coefficients $c_k(f)$ de (37), **pourvu que la série (45) converge vers f en tout point de $[-1, 1]$.**

On a vu plus haut des conditions suffisantes de convergence à partir de la théorie des séries de Fourier.

La raison de la restriction est que la série (45) peut en effet être divergente pour certaines fonctions continues. La raison profonde est que les opérateurs \mathcal{S}_n de (44) ne sont pas uniformément bornés (en n) dans $\mathcal{C}_{[-1,1]}$ muni de la norme du maximum¹⁵, d'où la possibilité de suites $\{\mathcal{S}_n(f)\}_n$ divergentes (conditions nécessaires du théorème de Banach-Steinhaus [et l'exemple ci-dessus]).

La détérioration de tout schéma de bonne approximation (choix de projecteurs linéaires) vis-à-vis des meilleurs approximations est quantifiée par l'énoncé suivant:

Lemme de Lebesgue. *Soit une famille $\{V_m\}_m$ de sous espaces vectoriels de l'espace vectoriel normé X , et $\forall m, \mathcal{P}_m$ un projecteur linéaire sur V_m . Alors, $\forall f \in X$, si $E_m(f) = \inf_{p \in V_m} \|f - p\|$, on a*

$$\|f - \mathcal{P}_m f\| \leq (1 + \|\mathcal{P}_m\|) E_m(f). \tag{46}$$

Ça ne sert à rien d'être le meilleur, ce qui est important c'est d'être bon (Christophe MALAVOY)

On appelle $\|\mathcal{P}_m\|$ la $m^{\text{ème}}$ **constante de Lebesgue** du schéma de bonne approximation $\{\mathcal{P}_m\}_m$. En effet, $\forall \varepsilon > 0$, prenons $p_\varepsilon \in V_m$ avec $\|f - p_\varepsilon\| \leq E_m(f) + \varepsilon$ [bien entendu, si V_m est de dimension finie, on prend plutôt une meilleure approximation \hat{p} de f]. Comme \mathcal{P}_m est un projecteur sur V_m , $\mathcal{P}_m p_\varepsilon = p_\varepsilon$ et, comme \mathcal{P}_m est linéaire,

$$\|f - \mathcal{P}_m f\| = \|f - \mathcal{P}_m f - (p_\varepsilon - \mathcal{P}_m p_\varepsilon)\| = \|f - p_\varepsilon - \mathcal{P}_m(f - p_\varepsilon)\| \leq (1 + \|\mathcal{P}_m\|)(E_m(f) + \varepsilon),$$

d'où (46). □

Ici, on trouve $\|\mathcal{S}_n\|_\infty = \frac{1}{\pi} \int_0^\pi \left| \frac{\sin(n+1/2)\varphi}{\sin \varphi/2} \right| d\varphi \rightarrow \infty$ quand $n \rightarrow \infty$, mais est bornée par $4 + (4/\pi^2) \ln n$ (si $n > 0$), ce qui est une croissance très lente:

0	1	2	3	4	5	6	7	8	9	10	20
1	1.436	1.642	1.778	1.880	1.961	2.029	2.087	2.138	2.183	2.223	2.494

Cf [Chen] pp. 127 & 149 (prob. 13), [Pas] p. 28, [Pow] pp. 143–149, [Riv1] p.61.

Remarque: si on connaît $\|f - \mathcal{S}_n f\|_\infty$, on a donc l'encadrement (assez médiocre) de $E_n(f)$:

$$\frac{\|f - \mathcal{S}_n f\|_\infty}{5 + (4/\pi^2) \ln n} \leq E_n(f) \leq \|f - \mathcal{S}_n f\|_\infty$$

¹⁵Par contre, les projecteurs \mathcal{S}_n sont uniformément bornés dans le même espace muni d'une norme quadratique appropriée, voir chapitre 3.

f	$c_k(f)$
$\frac{1-t^2}{2(1+t^2-2tx)}$	$t^k \quad (t < 1)$ (=fonction génératrice)
$\frac{1}{(1+t)^2-4tx^2}$	0 si k impair; $\frac{2t^{k/2}}{1-t^2}$ si k pair $(t < 1)$
$\frac{1}{x-c}$	$-\frac{4p^{k+1}}{1-p^2}$ avec $p = c \pm \sqrt{c^2-1}$ tel que $ p < 1 (c \notin [-1, 1])$
$ x $	0 si k impair; $-\frac{4(-1)^{k/2}}{\pi(k^2-1)}$ si k pair.
$\sqrt{1 \pm x}$	$-\frac{4\sqrt{2}(\mp 1)^k}{\pi(4k^2-1)}$
$\sqrt{1-x^2}$	0 si k impair; $-\frac{4}{\pi(k^2-1)}$ si k pair
$\arcsin x$	0 si k pair; $\frac{4}{\pi k^2}$ si k impair.
$\cos(q \arccos x)$	$-\frac{2q \sin(q\pi)(-1)^k}{\pi(k^2-q^2)}, \quad q \notin \mathbb{Z}$
$\ln(1-qx)$	$2 \ln\left(\frac{q}{2p}\right)$ si $k=0$; $-\frac{2p^k}{k}$ si $k > 0$, $\left(p = \frac{1-\sqrt{1-q^2}}{q}\right)$
$\arctg qx$	0 si k pair; $2(-1)^{(k-1)/2}\frac{p^k}{k}$ si k impair $\left(p = \frac{\sqrt{1+q^2}-1}{q}\right)$
e^{qx}	$2I_k(q)$ (fonction de Bessel $I_k(q) = i^{-k}J_k(iq) = \sum_{\ell=0}^{\infty} \frac{q^{k+2\ell}}{2^{k+2\ell}\ell!(k+\ell)!}$)
$J_\ell(qx)$	0 si $k-\ell$ impair; $2J_{(\ell-k)/2}(q/2)J_{(\ell+k)/2}(q/2)$ si $k-\ell$ pair

Quelques développements en série de Tchebycheff, S. Paszkowski [Pas].

On “retrouve” alors les formules (37) des coefficients par une exploitation (un peu cavalière) de la série (45): on multiplie les deux membres de (45) par $(1-x^2)^{-1/2}T_m(x)$ et on intègre sur $[-1, 1]$:

$$\int_{-1}^1 f(x) \frac{T_m(x) dx}{\sqrt{1-x^2}} = \sum_{k=0}^{\infty} c_k(f) \int_{-1}^1 \frac{T_k(x) T_m(x) dx}{\sqrt{1-x^2}} = \frac{\pi}{2} \sum_{k=0}^{\infty} c_k(f) \delta_{k,m} = \frac{\pi}{2} c_m(f).$$

La conséquence beaucoup plus intéressante et profonde des relations d’orthogonalité (36) est que la somme partielle de degré n de la série (45) est la meilleure approximation de f dans \mathcal{P}_n au sens d’une norme quadratique:

Proposition. La somme partielle $S_n := \mathcal{S}_n f = \sum_{k=0}^n c_k(f) T_k$ minimise $\int_{-1}^1 (f(x) - p(x))^2 \frac{dx}{\sqrt{1-x^2}}$

sur tous les $p \in \mathcal{P}_n$.

En effet, soit $p = S_n + q$, avec $q \in \mathcal{P}_n$. On a

$$\int_{-1}^1 \frac{(f(x) - p(x))^2 dx}{\sqrt{1-x^2}} = \int_{-1}^1 \frac{(f(x) - S_n(x))^2 dx}{\sqrt{1-x^2}} - 2 \underbrace{\int_{-1}^1 \frac{(f(x) - S_n(x))q(x) dx}{\sqrt{1-x^2}}}_0 + \int_{-1}^1 \frac{(q(x))^2 dx}{\sqrt{1-x^2}}$$

puisque l’intégrale de f fois $(1-x^2)^{-1/2}T_m(x)dx$ se confond avec celle de S_n fois $(1-x^2)^{-1/2}T_m(x)dx$ (on développe q dans la base des T_m). \square

Les meilleures approximations polynomiales de fonctions construites sur des principes d’orthogonalité sont si importantes que tout le chapitre 3 leur sera consacré.

Sommes de Fejér et de La Vallée Poussin.

Soit $S_n = \mathcal{S}_n f = \sum_0^n c_k(f) T_k$ qui peuvent donc ne pas converger vers f . La moyenne de Fejér

$$F_n := \frac{S_0 + S_1 + \cdots + S_n}{n+1}$$

converge uniformément vers f , mais la convergence est toujours médiocre.

Les sommes de la Vallée Poussin

$$V_n := \frac{S_n + S_{n+1} + \cdots + S_{2n-1}}{n} = 2F_{2n-1} - F_{n-1}$$

sont dans \mathcal{P}_{2n-1} au lieu de \mathcal{P}_n , mais ont des erreurs d’approximation qui se comparent aux erreurs de meilleure approximation: $\|f - V_n\|_{\infty} \leq 4E_n(f)$ ([DeVLor] pp. 273-274).

4.5. Vitesse de décroissance des coefficients et bornes de norme de fonction d’erreur.

Voici comment quantifier les erreurs de bonne approximation:

Théorème. Si $f \in \mathcal{C}_{[-1,1]}^m$,

$$|c_k(f)| \leq \frac{2\|F^{(m)}\|_{\infty}}{k^m} \quad (k > 0),$$

$$\|f - \mathcal{S}_n f\|_{\infty} \leq \frac{2\|F^{(m)}\|_{\infty}}{(m-1)n^{m-1}} \quad (n > 0, m > 1), \quad (47)$$

où $F(\theta) = f(\cos \theta)$. On a aussi

$$|c_k(f)| \leq \frac{4\|f^{(m)}\|_\infty}{\pi(k-m+1)(k-m+3)\dots(k+m-1)}, \quad (k \geq m \geq 0),$$

$$\|f - \mathcal{S}_n f\|_\infty \leq \frac{4\|f^{(m)}\|_\infty}{\pi(m-1)(n-m+2)(n-m+4)\dots(n+m-2)}, \quad (n \geq m, m > 1).$$

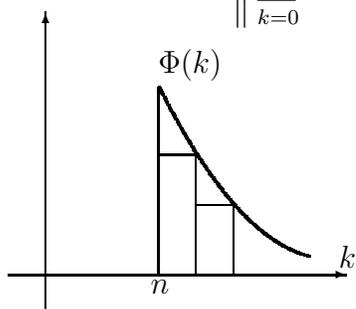
En effet, on intègre (37) par parties (l'intégrale en θ)

$$c_k(f) = \frac{2}{\pi} \left[F(\theta) \frac{\sin k\theta}{k} \right]_0^\pi - \frac{2}{\pi} \int_0^\pi F'(\theta) \frac{\sin k\theta}{k} d\theta,$$

les termes aux limites s'annulent, grâce au sinus. Si $f \in \mathcal{C}^2$, une nouvelle intégration par parties fournit des termes aux limites $F'(\theta) \cos k\theta$ en 0 et π qui s'annulent cette fois parce que $F : F(\theta) = f(\cos \theta)$ est une fonction périodique paire: $F(\theta) = F(-\theta) \Rightarrow F'(0) = 0$, $F(\pi + u) = F(-\pi + u) = F(\pi - u) \Rightarrow F'(\pi) = 0$. En poursuivant les intégrations par parties, on a des termes aux limites contenant soit des sinus, soit des dérivées d'ordre impair de F en 0 et π , ces dérivées sont nulles (considérer les développements de Taylor de F autour de 0 et π). Il reste $2/\pi$ fois une intégrale de $F^{(m)}(\theta)$ fois un sinus ou un cosinus divisé par k^m , d'où la première borne de $c_k(f)$ dans (47).

Pour la borne de $\|f - \mathcal{S}_n f\|$, on part de

$$\|f - \mathcal{S}_n f\|_\infty = \left\| \sum_{k=0}^\infty c_k(f)T_k - \sum_{k=0}^n c_k(f)T_k \right\|_\infty = \left\| \sum_{k=n+1}^\infty c_k(f)T_k \right\|_\infty \leq \sum_{k=n+1}^\infty |c_k(f)|,$$



et on utilise un principe très simplement illustré par la figure ci-contre: si Φ est positive décroissante,

$$\sum_{k=n+1}^\infty \Phi(k) \leq \int_n^\infty \Phi(k) dk$$

appliqué ici à $\Phi(k) = \text{constante}/k^m$. Bien entendu, l'intégrale ne converge que si $m > 1$.

Pour établir les autres inégalités, on part de l'intégrale en x de (37), que l'on intègre par parties, en dérivant m fois f et en intégrant m fois $V_{n,0}(x) = (1-x^2)^{-1/2}T_n(x)$, donc,

$$c_k(f) = (2/\pi)(-1)^m \int_{-1}^1 f^{(m)}(x)V_{n,m}(x) dx, \text{ d'après la définition (32), et on applique (33).} \quad \square$$

Exemple. Avec $f(x) = e^{qx}$, on obtient $\|f - \mathcal{S}_n f\|_\infty \leq 4q^n/(\pi 2^{n-1}(n-1)(n-1)!)$, (faire $m = n$ dans (47)), ce qui est intéressant: c'est presque 2^n fois plus petit que ce qu'on obtient avec la série de Taylor-Maclaurin¹⁶!

Autre exemple. Avec $f(x) = |x|$, on a *exactement* $\|f - \mathcal{S}_n f\|_\infty = 2/(\pi(n'-1))$, où n' est le plus petit nombre pair strictement plus grand que n : utiliser la table de la p. 59 et $\sum_{k>n} |c_k| = (4/\pi)[(n'^2-1)^{-1} + ((n'+2)^2-1)^{-1} + \dots] = (2/\pi)[1/(n'-1) - 1/(n'+1) + 1/(n'+1) - 1/(n'+3) + \dots]$, atteint en $x = 0$. remarquons que ce f n'est même pas dans \mathcal{C}^1 .

Approximation polynomiale de \sqrt{x} sur $[0,1]$.

Proposition. $\exists p \in \mathcal{P}_n : |\sqrt{x} - p(x)| \leq \frac{1}{\pi(n+1/2)}, 0 \leq x \leq 1.$

¹⁶De plus, (47) rend encore des services quand on ne dispose pas de coefficients de Taylor d'ordre élevé! (garder m faible dans (47)).

En effet, il suffit de prendre une somme partielle de la série de Tchebycheff de la fonction valeur absolue, en \sqrt{x} :

$$\left| \sqrt{x} + \frac{4}{\pi} \sum_0^n \frac{(-1)^k T_{2k}(\sqrt{x})}{4k^2 - 1} \right| \leq \frac{4}{\pi} \sum_{n+1}^{\infty} \frac{1}{4k^2 - 1} = \frac{2}{\pi} \sum_{n+1}^{\infty} \left(\frac{1}{2k - 1} - \frac{1}{2k + 1} \right) = \frac{2}{\pi(2n + 1)}.$$

Nous sommes maintenant en mesure d'aborder une première démonstration d'un théorème fondamental¹⁷

4.6. Théorème de Weierstrass. *Toute fonction f continue sur un intervalle compact $[a, b]$ peut être arbitrairement bien approchée en norme du maximum par des polynômes:*

$$\forall \varepsilon > 0, \exists n \text{ et } p \in \mathcal{P}_n : \|f - p\|_{\infty} \leq \varepsilon.$$

Définition. On appelle **module de continuité** d'une fonction f définie sur $[a, b]$ la **fonction**

$$\omega_f(h) := \max_{\substack{x, y \in [a, b] \\ |x - y| \leq h}} |f(x) - f(y)|, \quad 0 \leq h \leq b - a$$

Cette fonction est croissante, et $\omega_f(h) \rightarrow 0$ quand $h \rightarrow 0$ si f est continue sur $[a, b]$ borné (continuité uniforme de f). Et on montre aussi que ω_f est continue si f l'est, et $x, y \geq 0 \Rightarrow \omega_f(x + y) \leq \omega_f(x) + \omega_f(y)$. Donc aussi $\omega_f(kx) \leq k\omega_f(x)$ si k entier ≥ 0 ; enfin,

$$\omega_f(ax) \leq (a + 1)\omega_f(x), \tag{48}$$

si a est réel ≥ 0 (prendre $k =$ partie entière par excès de a).

Soit $h = (b - a)/N$ tel que $\omega_f(h) \leq \varepsilon/2$, g l'interpolant linéaire par morceaux de f aux points $x_i = a + ih$, $i = 0, 1, \dots, N$.

Entre x_i et x_{i+1} , $f(x) - g(x) = f(x) - \frac{(x_{i+1} - x)f(x_i) + (x - x_i)f(x_{i+1}))}{h} = \frac{(x_{i+1} - x)(f(x) - f(x_i)) + (x - x_i)(f(x) - f(x_{i+1}))}{h}$, donc $\|f - g\|_{\infty} \leq \omega_f(h) \leq \varepsilon/2$.

Abordons maintenant $g(x) - (\mathcal{S}_n g)(x) = \sum_{n+1}^{\infty} c_k T_k(y)$, où $x = (a + b)/2 + y(b - a)/2 \iff y = (2x - a - b)/(b - a)$.

$$\begin{aligned} c_k &= \frac{2}{\pi} \int_{-1}^1 g((a + b)/2 + y(b - a)/2) \frac{T_k(y)}{\sqrt{1 - y^2}} dy \\ &= -\frac{b - a}{\pi} \int_{-1}^1 g'((a + b)/2 + y(b - a)/2) V_{k,1}(y) dy \\ &= -\frac{b - a}{\pi} \sum_{i=0}^{N-1} g'(x_i) [V_{k,2}(y_{i+1}) - V_{k,2}(y_i)], \end{aligned}$$

et, comme $\|V_{k,2}\|_{\infty} \leq (k^2 - 1)^{-1}$ (par (33)), et que $|g'(x_i)| = |f(x_{i+1}) - f(x_i)|/h \leq \omega_f(h)/h$, $|c_k| \leq \frac{b - a}{\pi} N \frac{\omega_f(h)}{h} \frac{2}{k^2 - 1} = \frac{(b - a)^2 \omega_f(h)}{\pi h^2} \left(\frac{1}{k - 1} - \frac{1}{k + 1} \right)$, $\sum_{k=n+1}^{\infty} |c_k| \leq \frac{(b - a)^2 \omega_f(h)}{\pi h^2 n}$, et

$$\|f - \mathcal{S}_n g\|_{\infty} \leq \|f - g\|_{\infty} + \|g - \mathcal{S}_n g\|_{\infty} \leq \omega_f(h) \left(1 + \frac{(b - a)^2}{\pi h^2 n} \right),$$

¹⁷Il s'agit de la preuve de Lebesgue, voir A. Pinkus, Weierstrass and approximation theory, *J. Approx. Theory* **107** (2000) 1-66.

et il suffit de prendre $n \geq (b-a)^2/(\pi h^2)$. □

4.7. Calcul des coefficients de Tchebycheff et autres algorithmes.

- (1) On peut évidemment avoir la chance d'avoir une intégrale explicite, cf. table p. 59.
- (2) Si on dispose de nombreuses valeurs numériques de f , on peut estimer $c_k(f)$, $k = 0, 1, \dots, N$ par les coefficients discrets $c_k^{(N)}$ de (40) (p. 52; le résultat du calcul (40) ne coïncide plus avec (37) si f n'est pas un polynôme, aussi note-t-on ici $c_k^{(N)}(f)$ la valeur de (40)).

Ces coefficients peuvent être très efficacement calculés par un *algorithme de transformée discrète rapide (FFT)*

(cf. `tchebfft.m`)

```
% tchebfft.m
%
%   coefficients de Tchebycheff d'une fonction de x, -1<= x <= 1
%
nomf=input('donnez la fonction, par exemple: 3.*x+4, etc.
           N.B. x doit pouvoir etre un vecteur ', 's');
n=1;while n>0,
  n=input('nombre de subdivisions? (stop si <=0 ');
  if n<=0,
    break;
  end;
  % valeurs de f:
  clear x;x=cos( pi*(0:n)/n );
  clear f;f=eval(nomf);
  % passage a 2*n
  clear g;g=[f(n+1:-1:1) f(2:n)];
  coef=cos(pi*(0:2*n-1)) .* fft(g) /n;
  coef(1:n+1),
end
```

Ainsi, avec $\exp(x)$, on obtient, selon N :

N	$c_0^{(N)}$	$c_1^{(N)}$	$c_2^{(N)}$	$c_3^{(N)}$	$c_4^{(N)}$	$c_5^{(N)}$	$c_6^{(N)}$	$c_7^{(N)}$
1	3.0862	2.3504						
2	2.5431	1.1752	0.5431					
3	2.5322	1.1309	0.2770	0.0887				
4	2.5321	1.1303	0.2715	0.0449	0.0109			
5	2.5321	1.1303	0.2715	0.0443	0.0055	0.0011		
6	2.5321	1.1303	0.2715	0.0443	0.0055	0.0005	0.0001	
7	2.5321	1.1303	0.2715	0.0443	0.0055	0.0005	0.0000	0.0000

On converge donc rapidement vers les c_k de l'exponentielle

k	c_k
0	2.53213175550402
1	1.13031820798497
2	0.27149533953408
3	0.04433684984866
4	0.00547424044209
5	0.00054292631191
6	0.00004497732295
7	0.00000319843646
8	0.00000019921248
9	0.00000001103677
10	0.00000000055059
11	0.00000000002498
12	0.00000000000104
13	0.00000000000004

(cf. [Clen]).

La discordance entre $c_k(f)$ et $c_k^{(N)}(f)$ est remarquablement bien suggérée par la formule suivante (“*Aliasing*”, “*pliage fréquentiel*”):

$$c_k^{(N)}(f) = c_k(f) + c_{2N-k}(f) + c_{2N+k}(f) + c_{4N-k}(f) + c_{4N+k}(f) + \dots \quad (49)$$

En effet, on entre la “vraie” série (45) dans (40):

$$\begin{aligned} c_k^{(N)}(f) &= \frac{2}{N} \sum_{\ell=0}^N \left[\sum_{k'=0}^{\infty} c_{k'}(f) T_{k'}(\cos(\ell\pi/N)) \right] T_k(\cos(\ell\pi/N)) \\ &= \frac{2}{N} \sum_{k'=0}^{\infty} c_{k'}(f) P_{k,k'}^{(N)} \quad \text{le } P_{k,k'}^{(N)} \text{ de (42)!} \\ &= \sum_{p=0}^{\infty} c_{k+2pN}(f) + \sum_{p=1}^{\infty} c_{2pN-k}(f). \end{aligned}$$

Cette formule (49) représente exactement la superposition des fréquences observée quand on échantillonne un signal: à force de chercher des termes équiocillants, le langage et les méthodes de la théorie de l’approximation se rapproche de ce qui se fait en théorie du signal (thèse de Hamming).

(3) *Traitement de la récurrence.*

On manipule des séries de Tchebycheff comme on manipule des séries de Taylor. Ainsi, multiplier une série de Tchebycheff par un polynôme revient à effectuer plusieurs fois une multiplication par x :

$$\begin{aligned} x \sum_{k=0}^{\infty} c_k(f) T_k(x) &= \sum_{k=0}^{\infty} c_k(f) x T_k(x) \\ &= \sum_{k=0}^{\infty} c_k(f) \frac{T_{k+1}(x) + T_{k-1}(x)}{2} = \frac{c_1(f)}{2} T_0 + \sum_{k=1}^{\infty} \frac{c_{k-1}(f) + c_{k+1}(f)}{2} T_k(x) \end{aligned}$$

par (12). D’où $c_k(xf) = (c_{k+1}(f) + c_{k-1}(f))/2$, $k = 0, 1, \dots$ (si on pose $c_{-1} = c_1$). On peut aussi considérer des multiplications et des divisions de séries.

(4) *Economisation–réarrangement de polynômes.*

- Economisation: on applique à la lettre le mécanisme de cascade décrit en section 4.2 à partir d'une approximation polynomiale connue.

Exemple: $p(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24}$ est une approximation de e^x d'erreur ≤ 0.01 sur $[-1, 1]$ (le premier terme négligé est $x^5/120$). On ne peut supprimer $x^4/24$ sous peine d'avoir une erreur très supérieure à 0.01. Cependant, en soustrayant le multiple approprié de T_4 , on garde une approximation de degré 3 d'erreur encore acceptable:

$$p(x) - \frac{1}{192}T_4(x) = p(x) - \frac{1 - 8x^2 + 8x^4}{192} = \frac{191}{192} + x + \frac{13x^2}{24} + \frac{x^3}{6}$$

ne fait qu'ajouter une erreur $\leq 1/192$ à l'erreur initiale.

- Réarrangement: on construit le développement d'un polynôme $p(x) = \sum_0^N a_k x^k$ dans la base $\{T_0, \dots, T_N\}$.

Algorithme (Hamming): on réarrange successivement les polynômes partiels du schéma de Horner $a_N, a_N x + a_{N-1}, (a_N x + a_{N-1})x + a_{N-2}$, etc. en traitant chaque multiplication par x selon le point 3 ci-dessus. Ainsi:

k	a_k	c_0	c_1	c_2	c_3	c_4
4	1/24	1/12				
3	1/6	1/3	1/24			
2	1/2	25/24	1/6	1/48		
1	1	13/6	51/96	1/12	1/96	
0	1	243/96	27/24	13/48	1/24	1/192

(5) *Relations et équations différentielles.*

- Méthode des τ (Lanczos).

L'introduction d'un polynôme p dans un opérateur différentiel $\mathcal{L}y$ (par exemple, $\mathcal{L}y = y' - y$), ne va normalement pas annuler l'opérateur ($p' - p$ est évidemment toujours du degré de p). On modifie le second membre par un polynôme de norme minimale (\Rightarrow pol. de Tchebycheff), de façon à rendre le problème soluble dans \mathcal{P}_N . Un ou plusieurs coefficients d'abord indéterminés (les fameux τ) sont précisés *in fine* par la comparaison avec des valeurs initiales ou aux limites.

Exemple: résoudre $y' - y = 0$, avec $y(0) = 1$, dans \mathcal{P}_4 : comme il faut un second membre $\in \mathcal{P}_4$ également, on résout plutôt

$$p' - p = \left(\sum_0^4 a_k x^k \right)' - \sum_0^4 a_k x^k = \tau T_4(x),$$

soit $a_1 - a_0 + (2a_2 - a_1)x + (3a_3 - a_2)x^2 + (4a_4 - a_3)x^3 - a_4x^4 = \tau - 8\tau x^2 + 8\tau x^4$, d'où $a_4 = -8\tau$, $a_3 = -32\tau$, $a_2 = -88\tau$, $a_1 = -176\tau$, $a_0 = -175\tau$. Condition initiale $a_0 = 1 \Rightarrow \tau = -1/175$, etc.

- Relations de récurrence entre coefficients (Clenshaw).

On résout un problème différentiel en substituant une série de Tchebycheff à la fonction inconnue. En plus des opérations mentionnées au 3 ci-dessus, il faut traiter les relations différentielles. Grâce à (30) (p. 50), on a une relation simple

entre une série de Tchebycheff et sa *primitive*:

$$f' = \sum_{k=0}^{\infty} c_k(f') T_k \Rightarrow f = \text{constante} + \frac{c_0(f')}{2} T_1 + \frac{c_1(f')}{2} T_2 + \sum_{k=2}^{\infty} c_k(f') \left[\frac{T_{k+1}}{2(k+1)} - \frac{T_{k-1}}{2(k-1)} \right],$$

donc, $c_k(f) = [c_{k-1}(f') - c_{k+1}(f')]/(2k)$, $k = 1, 2, \dots$. On obtient ainsi des relations de récurrence qu'il faut exploiter convenablement (cf. [FoxP]).

Par exemple, $f' = f$ avec $f(0) = 1$: $c_k = (c_{k-1} - c_{k+1})/(2k)$, $k = 1, 2, \dots$. Négligeons c_5 , $c_4 = c_3/8 \Rightarrow c_3 = 8c_4$, $c_3 = (c_2 - c_4)/6 \Rightarrow c_2 = 49c_4$, $c_2 = (c_1 - c_3)/4 \Rightarrow c_1 = 204c_4$, $c_1 = (c_0 - c_2)/2 \Rightarrow c_0 = 457c_4$. Condition initiale $c_0/2 + c_1 T_1(0) + c_2 T_2(0) + c_3 T_3(0) + c_4 T_4(0) = 1 \Rightarrow c_0/2 - c_2 + c_4 = 1 : c_4 = 2/361$, etc.

Nombreuses applications dans D. Gottlieb, S.A. Orszag: *Numerical Analysis of Spectral Methods: Theory and Applications*. SIAM (CBMS) 1977, C. Canuto, M.Y. Hussaini, A. Quarteroni, T. Zang: *Spectral Methods in Fluid Dynamics*, Springer, 1988, B. Fornberg: *A Practical Guide to Pseudospectral Methods*, Cambridge U.P., 1996, *Chebyshev Polynomials*, J. C. Mason and D. C. Handscomb, Chapman and Hall/CRC Press, 2003. Aussi le logiciel pseudopack, <http://www.cfm.brown.edu/people/wsd>

4.8. Algorithmes en représentation de Tchebycheff.

4.8.1. Algorithme de calcul de $S_n(x) = \sum_{k=0}^n c_k T_k(x)$: écrivons la récurrence pour $T_1(x), \dots, T_n(x)$

et faisons suivre de l'expression de $S_n(x)$:

$$\begin{bmatrix} x & -1 & & & & & \\ -1 & 2x & -1 & & & & \\ & & \dots & \dots & \dots & & \\ & & & \dots & \dots & \dots & \\ & & & & -1 & 2x & -1 \\ c_0/2 & c_1 & \dots & \dots & c_{n-1} & c_n & \end{bmatrix} \begin{bmatrix} T_0 \\ T_1(x) \\ T_2(x) \\ \vdots \\ T_{n-1}(x) \\ T_n(x) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ S_n(x) \end{bmatrix}$$

soit $\mathbf{Rt} = \boldsymbol{\sigma}$,

et multiplions par un vecteur ligne $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \dots, \alpha_n]$ tel que $\boldsymbol{\beta} = \boldsymbol{\alpha R}$ n'ait que sa seule première composante β_0 non nulle. On a alors $\beta_0 = \alpha_n S_n(x)$, donc $S_n(x) = \beta_0$ si on impose en plus $\alpha_n = 1$.

Calcul des α_k : $\alpha_{n+1} = 0$; $\alpha_n = 1$; $\alpha_{k-1} = 2x\alpha_k - \alpha_{k+1} + c_k$, $k = n, n-1, \dots, 1$.

Alors, $S_n(x) = \beta_0 = x\alpha_0 - \alpha_1 + c_0/2$.

4.8.2. *Zéros d'un polynôme*. x doit être tel que $S_n(x) = 0$, donc, le déterminant de la matrice ci-dessus doit être nul. Réduction à un problème de valeurs propres: on ajoute à l'avant-dernière ligne la dernière multipliée par $1/c_n$. On a

$$\begin{vmatrix} 2x & -2 & & & & & \\ -1 & 2x & -1 & & & & \\ & & \dots & \dots & \dots & & \\ & & & \dots & \dots & \dots & \\ c_0/(2c_n) & c_1/c_n & \dots & -1 + c_{n-2}/c_n & 2x + c_{n-1}/c_n & & \end{vmatrix} = 0.$$

Les racines de $S_n(x) = 0$ sont donc les valeurs propres de

$$\begin{bmatrix} 0 & 1 & & & & \\ 1/2 & 0 & 1/2 & & & \\ & \dots & \dots & \dots & & \\ & & \dots & \dots & \dots & \\ -c_0/(4c_n) & -c_1/(2c_n) & \dots & 1/2 - c_{n-2}/(2c_n) & -c_{n-1}/(2c_n) & \dots \end{bmatrix}.$$

(Méthode de Tchebycheff-Frobenius, voir Boyd, John P., Computing zeros on a real interval through Chebyshev expansion and polynomial rootfinding, *SIAM J. Numer. Anal.* **40** (2002), no. 5, 1666–1682; Stetter, Hans J., *Numerical polynomial algebra*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2004. xvi+472 pp; Day, David, Romero, Louis: Roots of polynomials expressed in terms of orthogonal polynomials. *SIAM J. Numer. Anal.* **43** (2005), no. 5, 1969–1987.

5. Approximation par fonction rationnelle.

Soit $\mathcal{R}_{m,n}$ l'ensemble des fonctions r pouvant se représenter comme $r = p/q$, avec $p \in \mathcal{P}_m$ et $q \in \mathcal{P}_n$. Si les degrés du numérateur et du dénominateur sont m' et n' lorsque r est réduite à sa plus simple expression, on dit que le défaut de r dans $\mathcal{R}_{m,n}$ est $\min(m - m', n - n')$.

On a alors le théorème d'équioscillation

Théorème. Toute fonction réelle f continue dans $[a, b]$ admet une seule meilleure approximation $\hat{r} \in \mathcal{R}_{m,n}$. La fonction d'erreur $f - \hat{r}$ admet un alternant d'au moins $m + n + 2 - \partial$ points, où ∂ est le défaut de \hat{r} . [Ach]

La réciproque est facile à montrer: si s était meilleure que \hat{r} , $s - \hat{r} = f - \hat{r} - (f - s)$ devrait changer au moins $m + n + 1 - \partial$ fois de signe dans (a, b) , mais $s - \hat{r} = \frac{p\hat{q} - \hat{p}q}{q\hat{q}}$ a un numérateur de degré au plus $m + n - \partial \dots$

Des algorithmes d'échange existent ([Ach, Pow, Riv1]), mais on a également développé des outils de nature plus analytique:

- (1) L'approximation de Padé¹⁸ $r_{m,n}^{(P)}$ reproduit un maximum de coefficients de Taylor, par exemple à l'origine. Normalement, $f(x) - r_{m,n}^{(P)}(x) = O(x^{m+n+1})$. Si $f(x) = \sum_{k=0}^{\infty} c_k x^k$, $r_{m,n}^{(P)}(x) = \frac{p(x)}{q(x)} = \frac{\sum_0^m p_k x^k}{\sum_0^n q_k x^k}$, où $[q_0, \dots, q_n]$ est une solution des n équations homogènes $\sum_{j=0}^n c_{m+k-j} q_j = 0, k = 1, \dots, n$, et $p =$ développement de Taylor de $f q$ limité au degré m .

Ainsi, si $f(x) = e^x$, $r_{0,0}^{(P)}(x) = 1, r_{1,0}^{(P)}(x) = 1 + x, r_{0,1}^{(P)}(x) = \frac{1}{1-x}, r_{1,1}^{(P)}(x) = \frac{1+x/2}{1-x/2}, r_{2,2}^{(P)}(x) = \frac{1+x/2+x^2/12}{1-x/2+x^2/12}$.

- (2) L'approximation de Padé-Tchebycheff¹⁹ $r_{m,n}^{(T)}$ étend l'idée précédente aux séries de polynômes de Tchebycheff:

$$f(x) = \sum_0^{\infty} c_k T_k(x) = \frac{c_0}{2} + \underbrace{\frac{1}{2} \sum_1^{\infty} c_k e^{ik\theta}}_{f_+(e^{i\theta})} + \underbrace{\frac{1}{2} \sum_1^{\infty} c_k e^{-ik\theta}}_{f_+(e^{-i\theta})}$$

et on prend $r_{m,n}^{(T)}(x) = \frac{c_0}{2} + \frac{r_{m,n}^{(P)}(e^{i\theta})}{2} + \frac{r_{m,n}^{(P)}(e^{-i\theta})}{2}$.

Le développement en série de Tchebycheff de $r_{m,n}^{(T)}$ coïncide normalement avec celui de f jusqu'au terme en T_{m+n} inclus. La fonction d'erreur est cependant souvent loin d'être proche de l'équioscillation.

¹⁸cf. Henri Eugène Padé, 1863-1953, cf. G.A. Baker, Jr., *Essentials of Padé Approximants*, Ac. Press, new York, 1975; G.A. Baker, Jr., P. Graves-Morris, *Padé Approximants*, 2 vol., Addison-Wesley, 1981.

¹⁹Baker & Graves-Morris, *op. cit.* vol. 2, pp. 56-63.

- (3) L'approximation CF. On utilise une construction de Carathéodory et Fejér²⁰ produisant une fonction d'erreur parfaitement équi-oscillante... mais qui n'est pas une fonction rationnelle! On projette ensuite dans $\mathcal{R}_{m,n}$.

La fonction $\tilde{r}_{m,n}^{(CF)}$ est une fonction méromorphe avec exactement n pôles dans $|z| > 1$, approchant

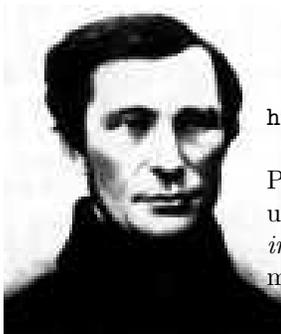
$$f_+^{(m,n)}(z) := \sum_{k=-n+1}^{\infty} c_k z^k \text{ telle que } f_+^{(m,n)}(z) - \tilde{r}_{m,n}^{(CF)}(z) = \sigma z^{m-n+1} \frac{\sum_{j=0}^{\infty} \bar{u}_j z^j}{\sum_{j=0}^{\infty} u_j z^{-j}} \text{ sur le cercle unité}$$

$|z| = 1$. σ est la $n^{\text{ème}}$ valeur singulière, et $[u_0, u_1, \dots]^T$ est le $n^{\text{ème}}$ vecteur singulier de la matrice de Hankel $[c_{i+j}]$, $i = m - n + 1, m - n + 2, \dots, j = 0, 1, \dots$ (matrice infinie ou tronquée à un ordre nettement supérieur à n).

Enfin, on projette $\sum_0^{m-n} c_k T_k(x) + \frac{\tilde{r}_{m,n}^{(CF)}(e^{i\theta}) + \tilde{r}_{m,n}^{(CF)}(e^{-i\theta})}{2}$ dans $\mathcal{R}_{m,n}$ (souvent par Padé-Tchebycheff).

Pour un programme `matlab` d'une simplicité et d'une efficacité confondantes, voir L.N. Trefethen, MATLAB programs for CF approximation, pp.599-602 in *Approximation Theory V*, (C.K.Chui, L.L.Schumaker, J.D.Ward, eds.), Academic Press, Orlando 1986, `cftref.m`

6. Lecture. Tchebycheff et de La Vallée Poussin



P.L. Tchebycheff

<http://www-history.mcs.st-and.ac.uk/history/BigPictures/Chebyshev.jpeg>

Pafnuty Lvovitch Tchebycheff (П.Л. Чебышев) (16 mai 1821–8 déc. 1894) fut un des esprits les plus remarquables du XIX^{ème} siècle. On connaît surtout son *inégalité de Tchebycheff* en probabilités: si X est une variable aléatoire réelle de moyenne m et de variance σ^2 ,

$$\text{Prob}(|X - m| \geq k\sigma) \leq \frac{1}{k^2}.$$

Cette inégalité se déduit simplement de ce que la fonction caractéristique de $\mathbb{R} \setminus [m - k\sigma, m + k\sigma]$ est bornée sur tout \mathbb{R} par $(x - m)^2 / (k\sigma)^2$, on aura l'occasion de revoir cette construction...

Il a également donné ce curieux et intéressant théorème d'analyse: la primitive de $x^m(a + bx^n)^p$ (*problème des différentielles binômes*) s'exprime au moyen de fonctions élémentaires si et seulement si l'un des trois nombres $\frac{m+1}{n}$, p , ou $\frac{m+1}{n} + p$ est entier...

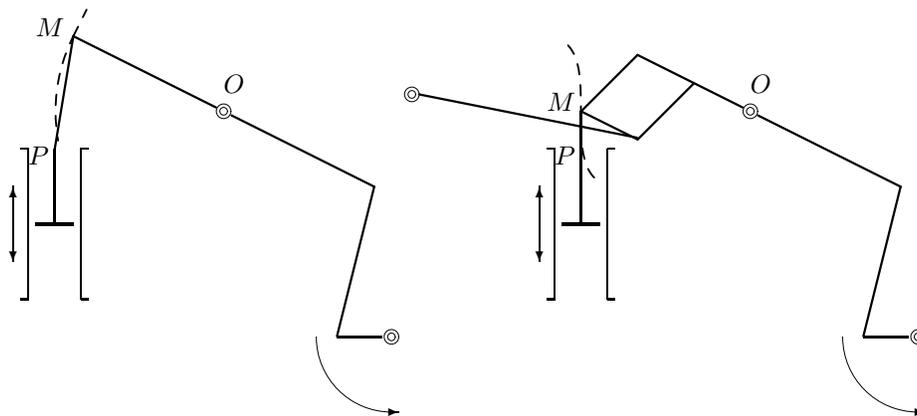
Il s'est également intéressé à des points fins de théorie des nombres: soit $\pi(x) :=$ nombre de nombres premiers $\leq x$. Comment se comporte $\pi(x)$ quand $x \rightarrow \infty$? Gauss avait conjecturé que $\lim_{x \rightarrow \infty} \frac{\pi(x)}{x / \ln x} = 1$, Tchebycheff

montra²¹ que les limites inférieure et supérieure de $\frac{\pi(x)}{x / \ln x}$ sont comprises entre 0.92129 et 1.1056. On reviendra sur ce résultat. Il donna aussi la première preuve de la conjecture de Bertrand: il y a toujours au moins un nombre premier entre n et $2n$ si $n > 3$.

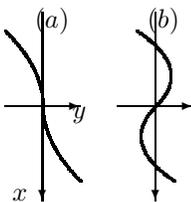
L'origine du problème de meilleure approximation résolu par Tchebycheff se situe dans la conception de mécanismes articulés associés à une machine à vapeur:

²⁰L.N. Trefethen, M. Gutknecht, The Carathéodory-Fejér method for real rational approximation, *SIAM J. Numer. Anal.* **20** (1983) 420-436.

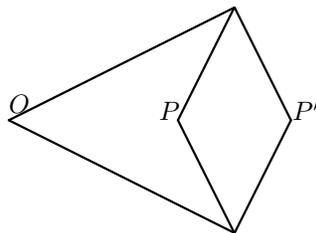
²¹W. Schwarz, Some remarks on the history of the prime number theorem from 1896 to 1960, pp.565–616 in J.P. Pier (ed.), *Development of Mathematics 1900–1950*, Birkhäuser, 1994.



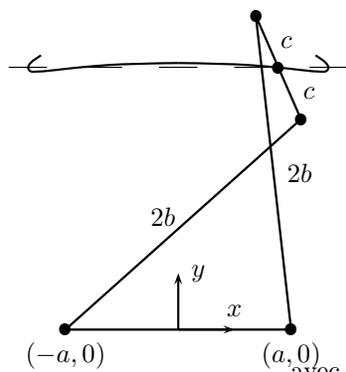
Mécanismes de transmission du mouvement.



Le mécanisme le plus simple (à gauche, dans la figure ci-dessus) force le point M à parcourir un arc de cercle, ce qui impose des efforts latéraux excessifs au piston P . Le mécanisme de droite, imaginé par James Watt, fait tracer au point M un arc de courbe possédant un point d'inflexion, se rapprochant ainsi nettement mieux de la verticale. Watt et ses successeurs se sont ingénier à trouver des mécanismes engendrant des points d'inflexion d'ordre de plus en plus élevé, avec une tangente verticale au point d'inflexion (figure (a)). Dans un voisinage $[-\delta, \delta]$ du point d'inflexion, l'équation de l'arc de courbe est proche de $y = Kx^n$ (l'axe des x est l'axe vertical). Tchebycheff montra que l'important n'est pas d'avoir un très bon contact avec la direction verticale en un point, mais bien de minimiser *le plus grand écart* par rapport à la ligne verticale idéale, ce qui l'amena au principe d'équioscillation (figure (b)).

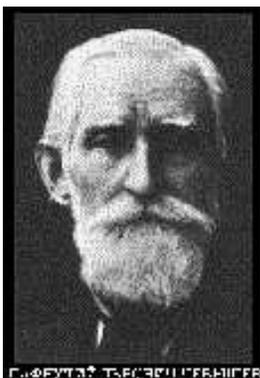


Ironie de l'histoire, un mécanisme transmettant de façon *exacte* un mouvement circulaire en un mouvement rectiligne (inverseur) fut imaginé par le général français Peaucellier en 1864... P et P' décrivent des figures inverses l'une de l'autre: $OP.OP' = \text{constante}$. Si P décrit un arc de cercle passant par O , P' décrit exactement un segment de droite. (Cf. § 16 de H. Rademacher, O. Toeplitz, *Von Zahlen und Figuren*, Springer, 1930, 1968 = *Enjoying Mathematics*)



Le mécanisme de Tchebycheff le plus connu (voir <http://ch.engr.ucdavis.edu/design/fourbar/fourbarChebyshev.html>; M. Nicaise, *Les mouvements mécaniques, étude descriptive et raisonnée des mécanismes*, Librairie polytechnique Ch. Béranger, Paris et Liège, 1931; D.C. Tao, *Applied Linkage Synthesis*, Addison-Wesley, Reading, 1964) est un quatre-barres symétrique de longueurs $2a$ pour la base, $2b$ pour les deux manivelles, et $2c$ pour la bielle. On s'intéresse à la trajectoire du point milieu (x, y) de la bielle. Soit c' la projection horizontale de la demi bielle: les coordonnées des extrémités de la bielle sont $(x \mp c', y \pm c')$, avec $c'^2 + c''^2 = c^2$, et exprimons que ces points sont à distance $2b$ de $(-a, 0)$ et $(a, 0)$: $4b^2 = (x - c' + a)^2 + (y + c'')^2 = (x + c' - a)^2 + (y - c'')^2$

avec $r^2 = x^2 + y^2$, on a $x^2 = \frac{r^2 c''^2}{4b^2 - r^2} = \frac{r^2(r^2 - 4b^2 + (a+c)^2)(4b^2 - (a-c)^2 - r^2)}{4b^2 - r^2}$,
 $y^2 = \frac{r^2(c' - a)^2}{4b^2 - r^2} - \frac{r^2(r^2 - 4b^2 - a^2 + c^2)^2}{4b^2 - r^2}$.. On voit que la courbe (réelle) est décrite quand $4b^2 - (a+c)^2 \leq r^2 \leq 4b^2 - (a-c)^2$. Un examen de la dérivée de y montre l'existence de minima de y en $x \neq 0$ si $(a+c)^3 < 4b^2 c$. Cf. [cheblink.m](#)



P.L. Tchebycheff (un peu plus âgé)

<http://www.math.psu.edu/dna/interpolation/interpolation.html>

Clenshaw ([Clen, p. 17]) fait une revue de quelques orthographes occidentales (alphabet latin) du nom de notre héros, les principales sont:

	Ч	E	Б	Ы	Ш	E	B
International	Č	E	B	Y	Š	E	V
Français	TCH	E	B	Y	CH	E	FF
Anglais	CH	E	B	Y	SH	E	V
Allemand	TSCH	E	B	Y	SCH	E	FF

On lira aussi P. Butzer, F. Jongmans: P.L. Chebyshev (1821–1894), a guide to his life and work, *J. Approx. Theory*, **96** (1999) 111-138; N.I. Akhiezer: Function theory according to Chebyshev, pp. 1-81 in A.N. Kolmogorov, A.P. Yushkevich (ed.): *Mathematics of the 19th Century*, vol. 3, Birkhäuser, 1998, V.L. Goncharov, The theory of best approximations of functions, *J. Approx. Theory* **106** (2000) 2-57; comments by V.M. Tikhomirov, *ibid.* 58-65.

<http://www-history.mcs.st-and.ac.uk/history/Mathematicians/Chebyshev.html>

History of Approximation Theory: <https://history-of-approximation-theory.com/>



C.J. de La Vallée Poussin, photo extraite de Butzer & Korevaar, (voir plus loin), aimablement numérisée par MM. Guy Buchet et Pierre Bulens.

Charles-Jean Etienne Gustave Nicolas, baron de La Vallée-Poussin (Louvain, 14 août 1866– Boitsfort, 2 mars 1962), professeur à l'université de Louvain, devint brusquement célèbre en 1896 par une preuve de la conjecture de Gauss relative au nombre $\pi(x)$ de nombres premiers $\leq x$: $\lim_{x \rightarrow \infty} \frac{\pi(x)}{x/\ln x} = 1$ ²². De plus, de La Vallée-Poussin confirma une autre intuition de Gauss, l'importance de la primitive de l'inverse du logarithme en cette matière:

$\left| \pi(x) - \int_2^x \frac{dt}{\ln t} \right| \leq \text{constante} \cdot x \cdot \exp(-\text{constante} \sqrt{\ln x})$. Riemann avait déjà poussé fort loin l'étude de la répartition des nombres premiers en exploitant l'identité sur la fonction zeta

$$\log \zeta(s) = \log \sum_{k=1}^{\infty} \frac{1}{k^s} = - \sum_p \log \left(1 - \frac{1}{p^s} \right) = \sum_p \left(\frac{1}{p^s} + \frac{1}{2p^{2s}} + \frac{1}{3p^{3s}} + \dots \right)$$

(voir p. 182 pour le prolongement analytique de $\zeta(s)$ à $\text{Re } s \leq 1$). Dès lors, $\frac{\log \zeta(s)}{s} = \mathcal{M} \pi(-s) + \mathcal{M} \pi(-2s) + \mathcal{M} \pi(-3s) + \mathcal{M} \pi(-4s) + \dots$, où \mathcal{M} est la transformée de Mellin $\mathcal{M} f(t) := \int_0^{\infty} f(x) x^{t-1} dx$, (remarquons que $\mathcal{M} f(kt)$ est la transformée de Mellin de $k^{-1} f(x^{1/k}) \dots$), ou encore $\mathcal{M} \pi(-s) = \frac{\log \zeta(s)}{s} - \frac{\log \zeta(2s)}{2s} - \frac{\log \zeta(3s)}{3s} - \frac{\log \zeta(5s)}{5s} + \frac{\log \zeta(6s)}{6s} + \dots$. En manipulant des formules d'inversion de la transformée de Mellin, Riemann arriva à $\pi(x) \sim R(x) - \frac{R(x^{1/2})}{2} - \frac{R(x^{1/3})}{3} - \frac{R(x^{1/5})}{5} + \frac{R(x^{1/6})}{6} + \dots$, avec $R(x) = \int_2^x \frac{dt}{\ln t} - \sum_{k=1}^{\infty} \int_2^{x^{\rho_k}} \frac{dt}{\ln t}$, où les ρ_k sont les zéros non réels de la fonction ζ^{23} . de La Vallée Poussin et Hadamard furent les premiers (c'est

²²Une autre démonstration fut donnée indépendamment la même année par Hadamard (1865– 1963 : les nombres premiers conservent).

²³D. Zagier, The first 50 million prime numbers, *The Math. Intelligencer* **0** (1977) 7-19; W. Schwarz, *op. cit.*

le mot) à montrer rigoureusement que la contribution de ces zéros est négligeable dans le comportement asymptotique de $R(x)$ quand $x \rightarrow \infty$. Riemann se proposait de montrer que tous ces zéros ont une partie réelle égale à $1/2$, mais la preuve n'est toujours pas connue (on cherche...) Cf. The encoding of the prime distribution by the zeta zeros (elegant approach) <http://www.maths.ex.ac.uk/~mwatkins/zeta/encoding2.htm>

de La Vallée Poussin rédigea de nombreux mémoires [dLVP2] et livres sur les relations entre sommes partielles de Fourier (ou de Tchebycheff) et meilleures approximations²⁴, ainsi que sur des questions fondamentales de théorie des fonctions, de la mesure et du potentiel. Nul doute que ces derniers travaux furent inspirés par les premiers.

n	$n \parallel x - \hat{p}_n(x) \parallel_\infty$
2	0.25000 00000 00000 00000
4	0.27048 35971 11137 10107
6	0.27557 43724 01175 38604
8	0.27751 78246 75052 69646
10	0.27845 11855 35508 60152
20	0.27973 24337 71973 82968
30	0.27997 46066 86407 49231
40	0.28005 97447 60423 15265
50	0.28009 92184 52382 83558
100	0.28015 19162 35465 27355

Dans un article délicieusement intitulé “Sur les polynômes d’approximation et la représentation approchée d’un angle” (*Bull. Acad. Roy. Belg.* **12** (1910) 808–844 = [dLVP2] VI 155–191), notre auteur explore la meilleure approximation polynomiale de degré n de $|x|$ sur $[-1, 1]$. Il obtient l’encadrement $\frac{A}{n} \leq E_n(|x|) \leq \frac{B}{n}$, la borne inférieure étant la plus difficile à établir (c’est là que de La Vallée Poussin introduit son théorème (§ 2.2, p. 31)). Plus tard, S. Bernstein (“Sur la meilleure approximation de $|x|$ par des polynômes de degrés donnés”, *Acta Math.* **37** (1914) 1–57) put établir l’existence

de la limite $C = \lim_{n \rightarrow \infty} nE_n(|x|)$, limite dont la valeur est encore aujourd’hui une énigme (cf. chap. 1 de R.S. Varga, *Scientific Computation on Mathematical Problems and Conjectures*, SIAM (CBMS-NSF **60**), 1990; M.I. Ganzburg, The Bernstein constant and polynomial interpolation at the Chebyshev nodes, *J. Approx. Th.* **119** (2002) 193–213; aussi l’article de Lubinsky cité en p. 40), et On the Bernstein Constants of Polynomial Approximation, par le même, dans *Constr. Approx.* (2007) **25**: 303–366 ...

(cf. constants.txt)

URL: <http://www.cecm.sfu.ca/projects/ISC/I.d.html>

BASE TABLE OF CONSTANTS

By Simon Plouffe , CECM, Centre for Experimental \& Constructive Mathematics \\
Simon Fraser University , Last update : Feb 6, 1996.

...

Constants associated with the Approximation of Functions

- * Wilbraham-Gibbs constant
- * Favard constants
- * The "one-ninth" constant
- * Lebesgue constants
- * Bernstein's constant

The Berstein Constant.

0.28016949902386913303643649123067200004248213981236

URL=<http://www.cecm.sfu.ca/projects/ISC/dataB/isc/C/bernstein.txt>

Professeur pendant de nombreuses décennies à l’université, il y exposa son cours d’analyse, un classique, <http://www.archive.org/details/danalyseinfinitesi01pousrich> à en juger par les nombreuses bibliothèques d’unité qui le conservent précieusement:

²⁴J. Favard, Hommage à Charles de La Vallée Poussin (1866–1962), pp. 1–3 in P.L. Butzer & J. Korevaar (éditeurs): *On Approximation Theory– Über Approximationstheorie*, Birkhäuser (ISNM **5**), 1964,1972. Voir aussi: J.C. Burkhil, Vallée Poussin, *J. London Math. Soc.* **39** (1964), 165–175. P.L. Butzer and R.J. Nessel, Aspects of de La Vallée Poussin’s Work in Approximation and its Influence, *Archive for History of Exact Science* **46** (1993–94), 67–95; <http://www-history.mcs.st-and.ac.uk/history/Mathematicians/ValleePoussin.html>; J. Mawhin, Charles-Jean de La Vallée Poussin, *Disquisitiones Mathematicæ* **1**, n° 1 (1998), 2–4, http://gauss.math.ucl.ac.be/PagesMath/Bienvenue/de_la_vallee_Bio.html

	Resultats du balayage	Nombre de Resultats
Getting http://www.bib.ucl.ac.be	De La Vallee Poussin	1
	De La Vallee Poussin, C	3
Libellule.	De La Vallee Poussin, Catherine	2
	De La Vallee Poussin, Ch.I	2
Catalogue	De La Vallee Poussin, Ch.-J	9
	De La Vallee Poussin, Ch.J	25
contenu	De La Vallee Poussin, Charles	2
Accès au catalogue	De La Vallee Poussin, Charles J	2
	de la Vallee Poussin, Charles Jean	10
	De La Vallee Poussin, Dominique	1
	De La Vallee Poussin,Et	1
	...	
	DE LA VALLEE POUSSIN, CH.J.	
B010075U	COURS D'ANALYSE INFINITESIMALE 1	LOUVAIN: UYSTPRUYST, 1903.-- 386
B010076N	COURS D'ANALYSE INFINITESIMALE 2	LOUVAIN: UYSTPRUYST, 1906.-- 456
B010073H	COURS D'ANALYSE INFINITESIMALE 1. 2e éd.	LOUVAIN: UYSTPRUYST, 1909.-- 423
...		
U3LABL1270	COURS D'ANALYSE INFINITESIMALE SIMONART, FERNAND, collab.	PARIS: GAUTHIER-VILLARS, 1926.-- 436
U1PCES0883	COURS D'ANALYSE INFINITESIMALE. 10e éd.	LOUVAIN: UYSTPRUYST, 1947.-- 481
U3ELHY0291	COURS D'ANALYSE INFINITESIMALE. 10e éd.	LOUVAIN: UYSTPRUYST, 1947.-- 489
U3ELHY0292	COURS D'ANALYSE INFINITESIMALE. 8e éd.	LOUVAIN: UYSTPRUYST, 1949.-- 556
U3FORT0551	COURS D'ANALYSE INFINITESIMALE. 8e éd.	LOUVAIN: UYSTPRUYST, 1949.-- 556
U3FORT0550	COURS D'ANALYSE INFINITESIMALE. 11e éd.	LOUVAIN: UYSTPRUYST, 1954.-- 492
U1PCES0864	COURS D'ANALYSE INFINITESIMALE. 9e éd.	LOUVAIN: UYSTPRUYST, 1957.-- 560
U3PHYS0346	COURS D'ANALYSE INFINITESIMALE. 9e éd.	LOUVAIN: LIBR.UNIVERSITAIRE, 1957.-- 560
U1PCES0884	COURS D'ANALYSE INFINITESIMALE. 12e éd.	LOUVAIN: LIBRAIRIE UNIVERSITAIRE, 1959.-- 492
U3PHYS0345	COURS D'ANALYSE INFINITESIMALE. 12e éd.	LOUVAIN: LIBR.UNIVERSITAIRE, 1959.-- 490
	Le cours d'analyse de La Vallée Poussin à l'UCL	
B301430B	ETUDE DES INTEGRALES A LIMITES INFINIES POUR LESQUELLES LA FONCTION SOUS LE SIGNE EST CONTINUE	BRUXELLES: F.HAYEZ, 1892.-- 33
B301187Q	RECHERCHES ANALYTIQUES SUR LA THEORIE DES NOMBRES PREMIERS	BRUXELLES: HAYEZ, 1896
B301431U	SUR L'APPROXIMATION DES FONCTIONS D'UNE VARIABLE REELLE ET DE LEURS DERIVEES PAR DES POLYNOMES ET DES SUITES LIMITEES DE FOURIER	BRUXELLES: HAYEZ, 1908.-- 64
B304173A	INTEGRALES DE LEBESGUE/FONCTIONS D'ENSEMBLE /CLASSES DE BAIRE	PARIS: GAUTHIER-VILLARS, 1916.-- 162
...		
B010788X	LECONS DE MECANIQUE ANALYTIQUE.vol.:VECTEURS/ CINEMATIQUE/DYNAMIQUE DU POINT/STATIQUE	LOUVAIN: UYSTPRUYST, 1924.-- 288
B010789R	LECONS DE MECANIQUE ANALYTIQUE.vol.: DYNAMIQUE DES SYSTEMES/DYNAMIQUE DU CORPS SOLIDE/EQUATIONS DE LA MECANIQUE/MECANIQUE DES FLUIDES	LOUVAIN: UYSTPRUYST, 1925.-- 326
U1PCES0866	LECONS DE MECANIQUE ANALYTIQUE.vol.:T.02	LOUVAIN: UYSTPRUYST, 1925.-- 315
A306927P	LES NOUVELLES METHODES DE LA THEORIE DU POTENTIEL	PARIS: HERMANN, 1937.-- 46
A306721E	LE POTENTIEL LOGARITHMIQUE. BALAYAGE ET REPRESENTATION CONFORME	LOUVAIN: UYSTPRUYST, 1949.-- 464
	Autres ouvrages de La Vallée Poussin à l'UCL	



C.J. de La Vallée Poussin

(un peu plus jeune)

http://www-history.mcs.st-and.ac.uk/history/Mathematicians/Vallee_Poussin.html

CHAPITRE 3

Approximation en moyenne quadratique.

1. Produit scalaire, orthogonalité, espace préhilbertien.

On a vu que les polynômes de Tchebycheff permettent d'exprimer une bonne approximation de f dans \mathcal{P}_n par une formule explicite

$$f(x) \approx S_n(x) := \sum_{k=0}^{k=n} c_k(f) T_k(x), \quad \text{avec } c_k(f) = \frac{2}{\pi} \int_{-1}^1 f(x) T_k(x) \frac{dx}{\sqrt{1-x^2}}, \quad k = 0, 1, \dots$$

cette écriture du coefficient c_k découlant de la propriété d'**orthogonalité**

$$\int_{-1}^1 T_i(x) T_j(x) \frac{dx}{\sqrt{1-x^2}} = 0, \quad i \neq j.$$

1.1. Produits scalaires sur \mathcal{P}_n .

On se propose maintenant d'examiner d'autres possibilités d'orthogonalité, c'est-à-dire différentes formules de *produit scalaires sur des espaces de fonctions*. Soit w une fonction positive (≥ 0) et intégrable sur un intervalle (a, b) de $\overline{\mathbb{R}}$ ($-\infty \leq a < b \leq +\infty$) avec $\int_a^b w(x) dx > 0$ (fonction **densité** ou **poids**). De plus, si a et/ou b est infini, on demandera $\int_a^b x^{2n} w(x) du < \infty$. Montrons alors que

$$(f, g) := \int_a^b f(x)g(x) w(x) dx \tag{50}$$

est un produit scalaire sur \mathcal{P}_n :

Rappel. Produit scalaire

- Un produit scalaire sur un espace vectoriel X défini sur \mathbb{R} est une forme bilinéaire symétrique définie positive:

$$\forall f, g, h \in X, \forall \alpha, \beta \in \mathbb{R}, \quad (\alpha f + \beta g, h) = \alpha(f, h) + \beta(g, h),$$

$$\forall f, g \in X, \quad (g, f) = (f, g),$$

$$\forall f \in X, f \neq 0 \Rightarrow (f, f) > 0.$$

- Un produit scalaire sur un espace vectoriel X défini sur \mathbb{C} est une forme sesquilinéaire hermitienne définie positive:

$$\forall f, g, h \in X, \forall \alpha, \beta \in \mathbb{C}, \quad (\alpha f + \beta g, h) = \alpha(f, h) + \beta(g, h),$$

$$\forall f, g \in X, \quad (g, f) = \overline{(f, g)},$$

$$\forall f \in X, f \neq 0 \Rightarrow (f, f) > 0.$$

- Inégalité de Cauchy-Schwarz:

$$|(f, g)| \leq [(f, f)(g, g)]^{1/2},$$

l'égalité n'ayant lieu que si f et g sont dépendants ($f = 0$ ou $g = Cf$, C scalaire).

- $\|f\| = (f, f)^{1/2}$ est une norme sur X .

Bien entendu, deux éléments de X sont **orthogonaux** entre eux si $(f, g) = 0$.

On vérifie que (50) convient effectivement à \mathcal{P}_n :

- (1) Existence de l'intégrale. w étant intégrable et f continue, le problème ne se pose que si a et/ou b est infini. f et g étant des polynômes de degrés $\leq n$, (50) se distribue en intégrales de monômes $\int_a^b x^m w(x) dx$, (*moments* de w), $0 \leq m \leq 2n$. Soit $a_0 < -1$ (à considérer si $a = -\infty$) et $b_0 > 1$ (à considérer si $b = +\infty$). On a $\int_a^b = \int_a^{a_0} + \int_{a_0}^{b_0} + \int_{b_0}^b$. Les intégrales sur (a, a_0) et (b_0, b) de $|x^m|w(x)$ sont majorées par les intégrales de $x^{2n}w(x)$, qui existent par hypothèse.
- (2) Caractère défini positif. Si f est un polynôme non nul, il ne s'annule qu'en un nombre fini de points, on peut retirer du support \mathcal{S} de w (points x tels que $\int_{x-\varepsilon}^{x+\varepsilon} w(t) dt > 0, \forall \varepsilon > 0$) un ensemble \mathcal{S}_η de mesure suffisamment petite pour que $\int_{\mathcal{S} \setminus \mathcal{S}_\eta} w(t) dt$ soit encore strictement positive, et tel que f^2 soit $\geq \delta > 0$ sur $\mathcal{S} \setminus \mathcal{S}_\eta$:

$$\int_a^b f^2(t)w(t)dt = \int_{\mathcal{S}} f^2(t)w(t)dt \geq \int_{\mathcal{S} \setminus \mathcal{S}_\eta} f^2(t)w(t)dt \geq \delta^2 \int_{\mathcal{S} \setminus \mathcal{S}_\eta} w(t)dt > 0. \quad \square$$

Si f et g sont susceptibles de prendre des valeurs *complexes*, il faut étendre (50) à

$$(f, g) = \int_a^b f(x)\overline{g(x)}w(x) dx. \quad (50')$$

On peut aussi étudier une formule de produit scalaire directement inspirée du produit scalaire usuel sur \mathbb{R}^N : soit $x_1 < x_2 < \dots < x_N$ des points fixés de $[a, b]$, alors

$$(f, g) = \sum_{m=1}^{m=N} f(x_m)g(x_m) \quad (51)$$

est un produit scalaire (*produit scalaire discret*) parfaitement admissible, que l'on a donc simplement construit avec les vecteurs $(f(x_1), \dots, f(x_N))$ et $(g(x_1), \dots, g(x_N))$ de \mathbb{R}^N ! Notons quand même que (51) n'est défini positif que sur un espace vectoriel suffisamment petit pour ne pas contenir de fonction $\neq 0$ s'annulant en chaque x_m : (51) est défini positif sur \mathcal{P}_{N-1} mais pas sur \mathcal{P}_N , où on a $(f, f) = 0$ avec $f(x) = (x - x_1) \dots (x - x_N) \neq 0$. Des produits scalaires discrets de type (51) interviendront dans des *problèmes de moindres carrés*. Si on se donne $w_1, w_2, \dots, w_N > 0$, on peut définir la somme pondérée

$$(f, g) = \sum_{m=1}^{m=N} w_m f(x_m)g(x_m) \quad (52)$$

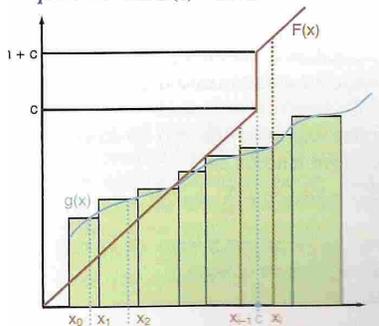
qui est encore un produit scalaire très utile.

La forme (52) commence à ressembler à (50)! L'intégrale (50) prise au sens de Riemann est effectivement une limite de sommes de type (52) lorsque les x_k se rapprochent indéfiniment.

La définition d'intégrale suivante permettra d'incorporer à la fois des intégrales et des sommes dans un produit scalaire:



Le mathématicien hollandais Thomas Stieltjes (1856-1894), naturalisé français en 1886, généralisa l'intégrale de Riemann à des fonctions définies sur un ensemble de points pondérés d'une masse : il répondait ainsi à une attente des physiciens, qui étaient souvent confrontés à ce genre de calcul. Il décrivit la répartition des masses ponctuelles au moyen d'une fonction F adéquate, puis, comme Riemann, subdivisa l'intervalle d'intégration en n parties $[x_{i-1}, x_i]$ et choisit, dans chacune, un point quelconque ξ_i . Il démontra que l'intégrale d'une fonction g sur la répartition de masses était la limite, quand n tend vers l'infini, de la somme $S_n = g(\xi_1)(F(x_1) - F(x_0)) + \dots + g(\xi_n)(F(x_n) - F(x_{n-1}))$. Cette intégrale est nommée intégrale de Stieltjes de g relativement à F . Dans son article sur les nombres premiers, Riemann avait été confronté à ce type d'intégrale, mais, dans le fil de son raisonnement, n'en avait tiré aucune loi générale. Ci-dessous, la fonction F décrit la présence d'une masse m au point c : $F(x) = x$ pour $x \neq c$ et $F(c) = m+c$.



Dans le même ouvrage, définition de l'intégrale de Lebesgue-Stieltjes aux §§ 56–58.

(in 'Riemann, les génies de la science', trimestriel *Pour La Science*, Août-Nov. 2002.)

Définition. Soit μ une fonction réelle bornée sur $[a, b]$ et f une fonction définie sur $[a, b]$. On appelle **intégrale de Riemann-Stieltjes**

$$\int_a^b f(x) d\mu(x) := \lim_{\max |x_m - x_{m-1}| \rightarrow 0} \sum_{m=1}^{m=M} f(x'_m) [\mu(x_m) - \mu(x_{m-1})]$$

si cette limite, prise sur les découpages $a = x_0 < x_1 < \dots < x_M = b$, $x_{m-1} \leq x'_m \leq x_m$, existe.

On démontre¹ que $\int_a^b f d\mu$ existe si a et b sont finis, f continue, et μ **croissante**² (pas nécessairement strictement croissante). Si a et/ou b est infini, on étudie la/les limites quand $a \rightarrow -\infty$ et/ou $b \rightarrow +\infty$. On voit donc que

$$(f, g) = \int_a^b f(x)g(x) d\mu(x) \tag{53}$$

est un produit scalaire généralisant (50) et (52) sur \mathcal{P}_n , pourvu que $\int_a^b x^{2n} d\mu(x) < \infty$, et que la fonction croissante μ possède au moins $n + 1$ points de croissance (points x tels que $\mu(x + \varepsilon) - \mu(x - \varepsilon) > 0, \forall \varepsilon > 0$).

Cette dernière précaution doit en effet être prise au cas où μ est une fonction en escalier, (53) donnant alors (52), les "marches" de hauteurs w_m de la fonction étant situées aux points x_m . Au contraire, si μ est (croissante) et dérivable sur (a, b) , (53) donne (50) avec $w = \mu'$.

En fait, on retrouve (50) dès que μ est *absolument continue*. Une fonction μ est absolument continue sur (a, b) si, $\forall \varepsilon > 0, \exists \delta > 0$: pour toute réunion $R = \bigcup [x_m, y_m] \subseteq (a, b)$ de longueur $\sum (y_m - x_m) \leq \delta$, on a $\sum |\mu(y_m) - \mu(x_m)| \leq \varepsilon$. Pour une fonction μ croissante, notons $\mu(R) = \sum (\mu(y_m) - \mu(x_m))$. On a alors longueur de $R \rightarrow 0 \Rightarrow \mu(R) \rightarrow 0$ si μ est absolument continue. Une fonction absolument continue sur (a, b) est toujours une primitive d'une fonction intégrable sur (a, b) : $\exists \nu \in L^1(a, b) : \mu(x) = \mu(a) + \int_a^x \nu(t) dt$.

On démontre que toute fonction croissante est la somme d'une fonction croissante absolument continue, d'une fonction en escalier (avec des "marches" éventuellement infiniment rapprochées), et d'une fonction croissante singulièrement continue... cf. par ex. Riesz & Sz.Nagy, *Leçons d'analyse fonctionnelle*, Gauthier-Villars, 5^{ème} éd., 1968, § 25.

¹P. Henrici, *Applied and Computational Complex Analysis*, II, Wiley, 1977, p.570. Il suffit de considérer $\sup \sum_m f_m [\mu(x_m) - \mu(x_{m-1})]$ et $\inf \sum_m F_m [\mu(x_m) - \mu(x_{m-1})]$, où f_m et F_m sont les valeurs minimales et maximales de f sur $[x_{m-1}, x_m]$.

²Interprétation probabiliste: si $\mu(a) = 0$ et $\mu(b) = 1$, on peut considérer μ comme une *fonction de répartition*.

Soit w positive et intégrable sur un **domaine D du plan complexe**, un produit scalaire valable est alors

$$(f, g) = \int_D f(z) \overline{g(z)} w(z) dx dy, \quad z = x + iy.$$

On peut aussi considérer

$$(f, g) = \int_C f(z) \overline{g(z)} w(z) ds, \quad ds = |dz|$$

sur une réunion d'arcs et de contours (rectifiables) du plan complexe.

Le produit scalaire **le plus général** sur \mathbb{R}^N a la forme $(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{M} \mathbf{y}$, où \mathbf{M} est une matrice carrée d'ordre N , symétrique définie positive (matrice de Gram, voir plus loin). La forme fonctionnelle correspondante est (pour des fonctions définies sur A) $(f, g) = \int_A \int_A K(x, y) f(x) g(y) dx dy$, avec $\forall f \in X, f \neq 0 : \int_A \int_A K(x, y) f(x) f(y) dx dy > 0$, plus généralement: $(f, g) = \int_A \int_A f(x) g(y) d\nu(x, y)$, où ν est une mesure sur $A \times A$ telle que $f \in X, f \neq 0 \Rightarrow (f, f) > 0$. On voit que (50) et (53) ne sont que des cas très particuliers où ν est concentrée sur le lieu $x = y$. Même en se limitant à $x = y$, on connaît encore des produits scalaires plus généraux que (53): ce sont les *produits scalaires de Sobolev* $(f, g) = \sum_{m=0}^M \int_A f^{(m)}(x) g^{(m)}(x) d\mu_m(x)$, très étudiés actuellement.

La spécificité des produits scalaires de type (53) sera spécialement mise à contribution à partir du § 3.2, c'est-à-dire que la plupart des énoncés qui seront donnés à partir du § 3.2 ne seront valables que pour des produits scalaires de type $\int_A f(x) g(x) d\mu(x)$ et non pas pour des produits scalaires plus généraux $\int_A \int_A f(x) g(y) d\nu(x, y)$, ni même pour des produits scalaires de type Sobolev.

1.2. Espace préhilbertien.

Définition. Un espace *préhilbertien* est un espace vectoriel muni d'un produit scalaire (\cdot, \cdot) et normé³ par $\|f\| = (f, f)^{1/2}$.

Inégalité de Cauchy-Schwarz: $(f, g) \leq \|f\| \|g\|$.

Identité du parallélogramme: $\|f + g\|^2 + \|f - g\|^2 = 2(\|f\|^2 + \|g\|^2)$.

2. Meilleure approximation dans un espace préhilbertien.

Dans un espace muni d'un produit scalaire (préhilbertien), le problème théorique de la détermination de meilleure approximation est infiniment plus simple que dans un espace normé général: on dispose maintenant d'une équerre en plus d'une règle graduée!

L'essentiel à retenir est que la détermination de la meilleure approximation dans un sous-espace vectoriel de dimension finie se réduit maintenant à la résolution d'un système d'équations *linéaires*, et que ce système est spécialement simple si on dispose d'une base *orthogonale* du sous-espace...

³N.B. Certains auteurs, dont L. Schwartz, Bourbaki, ou encore L. Chambadal et J.L. Ovaert dans leur article "Hilbert (espace de)" de l'*Encyclopedia Universalis* [1980], appellent préhilbertien un espace vectoriel muni d'une forme bilinéaire symétrique **semi**-définie positive, et donc susceptible de n'être que semi-normé. Il faut alors préciser "préhilbertien séparé", ou "euclidien" (ou "hermitien" pour les espaces sur \mathbb{C}), pour retrouver la définition utilisée ici. (Remarque communiquée par J. Meinguet).

2.1. Base d'un espace de dimension finie, matrice de Gram.

Considérons donc un sous-espace V de dimension finie d'un espace préhilbertien X . Soit $\{\mathbf{b}_0, \dots, \mathbf{b}_n\}$ une base de V (donc choisi de dimension $n + 1$). Tout élément de V s'exprime donc comme une combinaison linéaire unique des éléments de la base choisie:

$$f = \sum_{k=0}^n c_k \mathbf{b}_k.$$

La **matrice de Gram** de la base choisie relie les produits scalaires (f, \mathbf{b}_j) aux coefficients c_k :

$$[(f, \mathbf{b}_0) \quad \cdots \quad (f, \mathbf{b}_n)] = [c_0 \quad \cdots \quad c_n] \mathbf{G}_{n+1},$$

où la matrice de Gram \mathbf{G}_{n+1} est la matrice des produits scalaires $(\mathbf{b}_i, \mathbf{b}_j)$:

$$\mathbf{G}_{n+1} = \begin{bmatrix} (\mathbf{b}_0, \mathbf{b}_0) & \cdots & (\mathbf{b}_0, \mathbf{b}_n) \\ \vdots & \ddots & \vdots \\ (\mathbf{b}_n, \mathbf{b}_0) & \cdots & (\mathbf{b}_n, \mathbf{b}_n) \end{bmatrix}. \quad (54)$$

Cette matrice est symétrique (dans le cas d'un espace sur \mathbb{R}), hermitienne (dans le cas d'un espace sur \mathbb{C}), et **définie positive**:

Rappel. Matrice définie positive.

- Une matrice carrée \mathbf{A} réelle symétrique ($a_{i,j} = a_{j,i}$) est **définie positive** si

$$\forall \text{ vecteur réel } \mathbf{v} \neq 0, \quad \mathbf{v}^T \mathbf{A} \mathbf{v} > 0.$$

- Une matrice carrée \mathbf{A} complexe hermitienne ($a_{i,j} = \overline{a_{j,i}}$) est **définie positive** si

$$\forall \text{ vecteur complexe } \mathbf{v} \neq 0, \quad \mathbf{v}^H \mathbf{A} \mathbf{v} > 0,$$

(\mathbf{v}^H désigne le transposé conjugué de \mathbf{v} : $\mathbf{v}^H = \overline{\mathbf{v}^T}$).

- \mathbf{A} est alors nécessairement régulière (invertible): si le déterminant de \mathbf{A} était nul, il existerait un vecteur $\mathbf{v} \neq 0$ tel que $\mathbf{A} \mathbf{v} = 0$, donc $\mathbf{v}^H \mathbf{A} \mathbf{v} = 0$ impossible!
- On montre d'ailleurs que toutes les valeurs propres d'une telle matrice sont strictement positives, donc le déterminant = produit des valeurs propres > 0 .
- On dit encore que \mathbf{A} est **définie** si $\mathbf{v}^H \mathbf{A} \mathbf{v}$ a toujours le même signe dès que $\mathbf{v} \neq 0$, **non définie** si $\mathbf{v}^H \mathbf{A} \mathbf{v}$ s'annule pour au moins un vecteur $\mathbf{v} \neq 0$, **semi-définie positive** si $\mathbf{v}^H \mathbf{A} \mathbf{v} \geq 0$ pour tout \mathbf{v} , **indéfinie** si $\mathbf{v}^H \mathbf{A} \mathbf{v}$ prend des valeurs > 0 et < 0 .

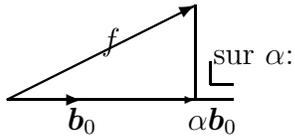
Montrons que \mathbf{G}_{n+1} est bien hermitienne définie positive si $\{\mathbf{b}_0, \dots, \mathbf{b}_n\}$ est bien une base de V : on a bien $(\mathbf{G}_{n+1})_{j,i} = (\mathbf{b}_j, \mathbf{b}_i) = \overline{(\mathbf{b}_i, \mathbf{b}_j)} = \overline{(\mathbf{G}_{n+1})_{i,j}}$ par symétrie hermitienne du produit scalaire; ensuite,

$$[c_0 \quad \cdots \quad c_n] \mathbf{G}_{n+1} \begin{bmatrix} c_0 \\ \vdots \\ \overline{c_n} \end{bmatrix} = [(f, \mathbf{b}_0) \quad \cdots \quad (f, \mathbf{b}_n)] \begin{bmatrix} c_0 \\ \vdots \\ \overline{c_n} \end{bmatrix} = (f, f) > 0$$

où $f = \sum_0^n c_k \mathbf{b}_k$, dès que $[c_0, \dots, c_n] \neq [0, \dots, 0]$, car f ne peut alors être nul.

2.2. Meilleure approximation = projection orthogonale.

Comme **exercice**, examinons d'abord le cas où V est de dimension... 1!



Soit $\mathbf{b}_0 \in V$, tous les éléments de V sont de la forme $\alpha\mathbf{b}_0$, il faut minimiser $\|f - \alpha\mathbf{b}_0\|$

$$\|f - \alpha\mathbf{b}_0\|^2 = (f - \alpha\mathbf{b}_0, f - \alpha\mathbf{b}_0) = (f, f) - \alpha(\mathbf{b}_0, f) - \overline{\alpha(\mathbf{b}_0, f)} + |\alpha|^2\|\mathbf{b}_0\|^2.$$

Si le champ des scalaires est \mathbb{R} , α est réel, on doit simplement minimiser un trinôme du second degré en α , et on trouve immédiatement $\alpha = (\mathbf{b}_0, f)/\|\mathbf{b}_0\|^2$.

Sur \mathbb{C} , voyons d'abord comment doit se comporter la phase (argument) de α si $|\alpha|$ est fixé:, $\alpha(\mathbf{b}_0, f) + \overline{\alpha(\mathbf{b}_0, f)}$ est maximal quand $\alpha(\mathbf{b}_0, f)$ est réel positif: $\alpha(\mathbf{b}_0, f) = |\alpha| |(\mathbf{b}_0, f)|$, il faut donc minimiser $\|f\|^2 - 2|\alpha| |(\mathbf{b}_0, f)| + |\alpha|^2\|\mathbf{b}_0\|^2$, soit $|\alpha| = |(\mathbf{b}_0, f)|/\|\mathbf{b}_0\|^2$, $\alpha = \overline{(\mathbf{b}_0, f)}/\|\mathbf{b}_0\|^2 = (f, \mathbf{b}_0)/\|\mathbf{b}_0\|^2$:

la meilleure approximation de la forme $\alpha\mathbf{b}_0$ de f est

$$\frac{(f, \mathbf{b}_0)}{\|\mathbf{b}_0\|^2} \mathbf{b}_0, \quad (55)$$

projection orthogonale de f dans V (voir dessin).

Passons maintenant à V de dimension finie quelconque:

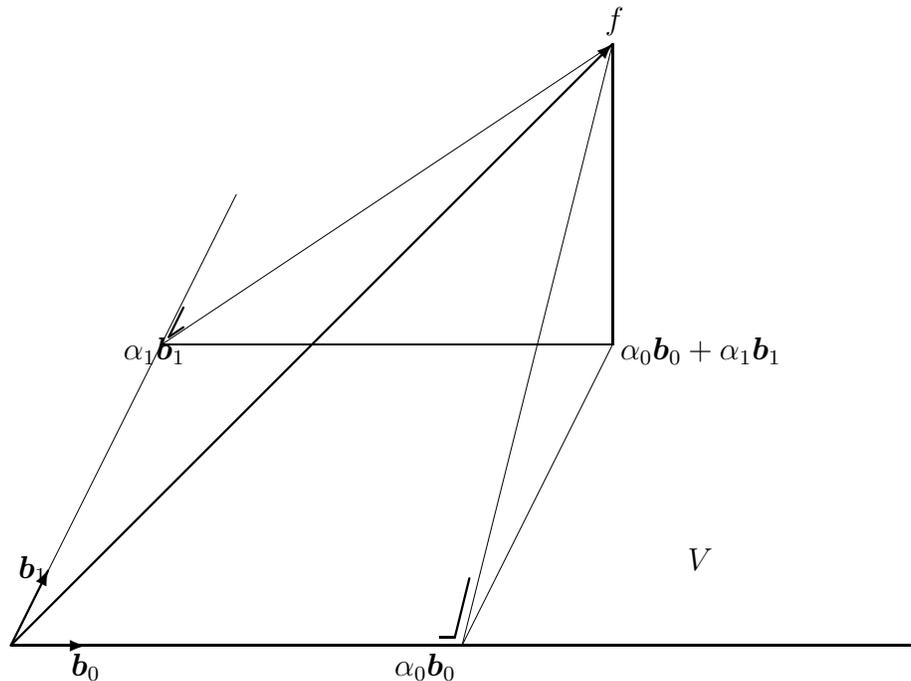


FIGURE 1. Projection orthogonale de f sur V muni d'une base orthogonale.

Théorème. Soit X un espace préhilbertien sur \mathbb{R} ou \mathbb{C} , de produit scalaire (\cdot, \cdot) , donc de norme $\|\cdot\| = (\cdot, \cdot)^{1/2}$, et V un sous-espace vectoriel de dimension finie de X .

Alors, $\forall f \in X$ a exactement une meilleure approximation $\hat{p} = Pf$ dans V ; cette meilleure approximation est l'application d'un projecteur linéaire sur X , le projecteur orthogonal $X \rightarrow V$: $f - \hat{p}$ est orthogonal à tout élément de V .

Si $\{\mathbf{b}_0, \dots, \mathbf{b}_n\}$ est une base de V , on obtient les coefficients du développement de $\hat{p} = Pf = \sum_{j=0}^{j=n} \alpha_j \mathbf{b}_j$ dans cette base en résolvant le système linéaire suivant (**équations normales**)

$$\mathbf{G}_{n+1}^T \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} (\mathbf{b}_0, \mathbf{b}_0) & \cdots & (\mathbf{b}_n, \mathbf{b}_0) \\ \cdots & \cdots & \cdots \\ (\mathbf{b}_0, \mathbf{b}_n) & \cdots & (\mathbf{b}_n, \mathbf{b}_n) \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} (f, \mathbf{b}_0) \\ \vdots \\ (f, \mathbf{b}_n) \end{bmatrix} \quad (56)$$

où on reconnaît la transposée de la matrice de Gram de la base de V .

En effet, construisons d'abord $\tilde{p} \in V$ tel que $f - \tilde{p}$ soit orthogonal à tout élément de V (en principe, on ne sait pas encore si ce \tilde{p} sera la meilleure approximation de f dans V): $f - \tilde{p} = f - \sum_{j=0}^{j=n} \tilde{\alpha}_j \mathbf{b}_j$ orthogonal à $\mathbf{b}_i, i = 0, 1, \dots, n$:

$$(f - \tilde{p}, \mathbf{b}_i) = (f, \mathbf{b}_i) - \sum_{j=0}^{j=n} \tilde{\alpha}_j (\mathbf{b}_j, \mathbf{b}_i) = 0, i = 0, \dots, n$$

ce qui est bien (56).

Comme nous savons que \mathbf{G}_{n+1} est définie positive, ce système a une solution unique, $\forall f \in X$. Montrons maintenant que $\tilde{p} = \hat{p}$, c'est-à-dire que l'orthogonalité du vecteur d'erreur au sous-espace d'approximation implique l'optimalité:

$$\forall p \in V, \quad \|f - p\|^2 = (f - \tilde{p} - (p - \tilde{p}), f - \tilde{p} - (p - \tilde{p})) = \|f - \tilde{p}\|^2 + \|p - \tilde{p}\|^2,$$

(**relation de Pythagore**), en effet, $(f - \tilde{p}, p - \tilde{p}) = (p - \tilde{p}, f - \tilde{p}) = 0$ par orthogonalité de $f - \tilde{p}$ et de tout élément de V . Donc, $\|f - p\| > \|f - \tilde{p}\|$ dès que $p \neq \tilde{p}$. \square

On peut expliciter le projecteur orthogonal sur V par

$$Pf = [\alpha_0 \quad \dots \quad \alpha_n] \begin{bmatrix} \mathbf{b}_0 \\ \vdots \\ \mathbf{b}_n \end{bmatrix} = [(f, \mathbf{b}_0) \quad \dots \quad (f, \mathbf{b}_n)] (\mathbf{G}_{n+1})^{-1} \begin{bmatrix} \mathbf{b}_0 \\ \vdots \\ \mathbf{b}_n \end{bmatrix}, \quad (57)$$

ce qui est un peu lourd. Les choses se simplifient remarquablement si on dispose d'une base orthogonale de V :

Si $\{\mathbf{b}_0, \dots, \mathbf{b}_n\}$ est une **base orthogonale** de V , la meilleure approximation de f dans V est donnée par

$$Pf = \sum_{j=0}^{j=n} \frac{(f, \mathbf{b}_j)}{\|\mathbf{b}_j\|^2} \mathbf{b}_j, \quad \text{si } (\mathbf{b}_i, \mathbf{b}_j) = 0, i \neq j. \quad (58)$$

En effet, la matrice de Gram \mathbf{G}_{n+1} est alors diagonale, d'éléments diagonaux $(\mathbf{b}_i, \mathbf{b}_i) = \|\mathbf{b}_i\|^2, i = 0, \dots, n$ dans ce cas.

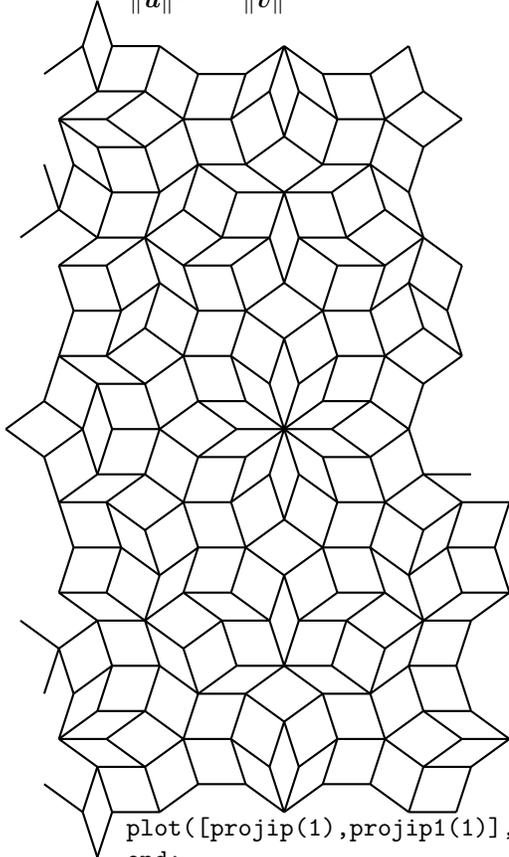
Cette formule (58) est donc à peine plus compliquée que (55)

La projection orthogonale sur un espace de dimension finie V n'est autre que la somme (vectorielle) des projections orthogonales sur les espaces de dimension un sous tendus par les éléments d'une base **orthogonale** de V .

Exemple. Projection orthogonale d'une partie de \mathbb{Z}^5 dans un plan (réseau quasi-périodique de de Bruijn-Penrose⁴).

Soit V le sous-espace de dimension 2 de \mathbb{R}^5 sous-tendu par les deux vecteurs (orthogonaux) \mathbf{u} et \mathbf{v} de composantes $\cos(k\pi/5)$ et $\sin(k\pi/5)$, $k = 1, \dots, 5$. A chaque point $\mathbf{p} = \alpha\mathbf{u} + \beta\mathbf{v}$ de V , on associe le point $\mathbf{q}(\mathbf{p})$ de \mathbb{Z}^5 le plus proche de \mathbf{p} (chaque composante de $\mathbf{q}(\mathbf{p})$ est l'entier le plus proche de la composante correspondante de \mathbf{p}). Enfin, on projette ce $\mathbf{q}(\mathbf{p})$ orthogonalement sur V tout simplement par (58):

$$\frac{\mathbf{u}'\mathbf{q}}{\|\mathbf{u}\|^2}\mathbf{u} + \frac{\mathbf{v}'\mathbf{q}}{\|\mathbf{v}\|^2}\mathbf{v}.$$



```

% de Bruijn-Penrose
% projection de Z5 sur le plan (u,v)
%
c25=sqrt(2.5);
u=cos((1:5)*pi/5)/c25;v=sin((1:5)*pi/5)/c25;
% check orthonormality in R5
[u*v',u*u',v*v'],
%
c=5;plot(-c:c,c*ones(size(-c:c)),'b.',-c:c,-c*ones(size(-c:c)),'b.');
```

```

ipv=[];ipvn=0;
h=0.02,ih=round(c/h);
for iu=-ih:ih, p2=-c;
while p2<c,
p=h*iu*u+p2*v;ip=round(p);iv=1;
ipvf=1;if ipvn==0, ipv=ip;ipvn=1;end;
ipd=ipv-ones(size(ipv,1),1)*ip;
ipd=abs(ipd)*[1 1 1 1 1]';
if min(ipd)==0,ipvf=0;end;
if ipvf>0, ipvn=ipvn+1;
ipv(ipvn,:)=ip;projip=[ip*u', ip*v'];
for j=1:ipvn-1;
projip1=[ipv(j,:)*u', ipv(j,:)*v'];
dt=norm(projip-projip1);
if abs(dt-0.400*c25)<0.01,
plot([projip(1),projip1(1)],[projip(2),projip1(2)]);
end;
end;
end;
end;
ipp2=ip-h*iu*u+0.51;p2=min( ipp2(1:4)./v(1:4) );
end; %iv
end; %u

```

2.3. Méthode d'orthogonalisation de Gram-Schmidt.

Le procédé de Gram-Schmidt consiste à construire progressivement une base orthogonale (dont on voit bien maintenant l'intérêt) à partir d'une base quelconque en retirant de chaque

⁴Voir American Mathematical Society :: Feature Column URL: <http://www.ams.org/featurecolumn/archive/penrose>
 Mathematics behind QuasiTiler by M. Senechal http://www.geom.uiuc.edu/apps/quasitiler/MS_about.html
 the Geometry Center <http://www.geom.uiuc.edu/>

élément de la base de départ une combinaison des éléments précédents de façon à être orthogonal à ces derniers: c'est une projection orthogonale, donc, d'après (58),

$$\begin{aligned} \mathbf{b}_0^\perp &= \mathbf{b}_0, \\ \mathbf{b}_1^\perp &= \mathbf{b}_1 - \frac{(\mathbf{b}_1, \mathbf{b}_0^\perp)}{\|\mathbf{b}_0^\perp\|^2} \mathbf{b}_0^\perp, \\ \mathbf{b}_2^\perp &= \mathbf{b}_2 - \frac{(\mathbf{b}_2, \mathbf{b}_0^\perp)}{\|\mathbf{b}_0^\perp\|^2} \mathbf{b}_0^\perp - \frac{(\mathbf{b}_2, \mathbf{b}_1^\perp)}{\|\mathbf{b}_1^\perp\|^2} \mathbf{b}_1^\perp, \\ &\text{etc.} \end{aligned}$$

Chaque nouvel élément d'une base orthogonale construite par le procédé de Gram-Schmidt est l'erreur de meilleure approximation de l'élément correspondant de la base initiale dans le sous-espace sous-tendu par les éléments précédents:

$$\mathbf{b}_m^\perp = \mathbf{b}_m - \sum_{k=0}^{m-1} \alpha_k^{(m)} \mathbf{b}_k^\perp, \text{ avec } \alpha_k^{(m)} = \frac{(\mathbf{b}_m, \mathbf{b}_k^\perp)}{\|\mathbf{b}_k^\perp\|^2} \quad (59)$$

2.4. Hauteurs, volumes et déterminants de Gram.

Si on développe l'inverse de la matrice de Gram dans (57), on obtient le développement de $\hat{p} = Pf$ dans la base des b_i , les coefficients étant des cofacteurs appropriés de \mathbf{G}_{n+1} divisés par le déterminant. Le résultat est

$$Pf = -\frac{1}{\det \mathbf{G}_{n+1}} \begin{vmatrix} (\mathbf{b}_0, \mathbf{b}_0) & \cdots & (\mathbf{b}_n, \mathbf{b}_0) & (f, \mathbf{b}_0) \\ \cdots & \cdots & \cdots & \cdots \\ (\mathbf{b}_0, \mathbf{b}_n) & \cdots & (\mathbf{b}_n, \mathbf{b}_n) & (f, \mathbf{b}_n) \\ \mathbf{b}_0 & \cdots & \mathbf{b}_n & 0 \end{vmatrix}$$

où le déterminant est interprété comme une combinaison linéaire des éléments (vectoriels) de la dernière ligne multipliés par les cofacteurs (purement scalaires, cette fois) appropriés; le produit scalaire d'une telle expression avec tout $g \in X$ est cette fois le "vrai" déterminant

$$-\frac{1}{\det \mathbf{G}_{n+1}} \begin{vmatrix} (\mathbf{b}_0, \mathbf{b}_0) & \cdots & (\mathbf{b}_n, \mathbf{b}_0) & (f, \mathbf{b}_0) \\ \cdots & \cdots & \cdots & \cdots \\ (\mathbf{b}_0, \mathbf{b}_n) & \cdots & (\mathbf{b}_n, \mathbf{b}_n) & (f, \mathbf{b}_n) \\ (\mathbf{b}_0, g) & \cdots & (\mathbf{b}_n, g) & 0 \end{vmatrix}.$$

Vérifions la formule pour Pf : le produit scalaire de Pf et n'importe quel b_i vaut (f, b_i) (par orthogonalité de $f - Pf$ et V). Or, avec $g = b_i$, et en soustrayant la $i^{\text{ème}}$ ligne de la dernière, il reste

$$-\frac{1}{\det \mathbf{G}_{n+1}} \begin{vmatrix} (\mathbf{b}_0, \mathbf{b}_0) & \cdots & (\mathbf{b}_n, \mathbf{b}_0) & (f, \mathbf{b}_0) \\ \cdots & \cdots & \cdots & \cdots \\ (\mathbf{b}_0, \mathbf{b}_n) & \cdots & (\mathbf{b}_n, \mathbf{b}_n) & (f, \mathbf{b}_n) \\ 0 & \cdots & 0 & -(f, b_i) \end{vmatrix},$$

ce qui vaut bien (f, b_i) .

Pour $f - Pf$, on a

$$f - Pf = \frac{1}{\det \mathbf{G}_{n+1}} \begin{vmatrix} (\mathbf{b}_0, \mathbf{b}_0) & \cdots & (\mathbf{b}_n, \mathbf{b}_0) & (f, \mathbf{b}_0) \\ \cdots & \cdots & \cdots & \cdots \\ (\mathbf{b}_0, \mathbf{b}_n) & \cdots & (\mathbf{b}_n, \mathbf{b}_n) & (f, \mathbf{b}_n) \\ \mathbf{b}_0 & \cdots & \mathbf{b}_n & f \end{vmatrix}.$$

Revenons à (56) et cherchons à déterminer la norme de l'erreur de meilleure approximation de f dans V . On a $\|f - \hat{p}\|^2 = (f - \hat{p}, f - \hat{p}) = (f - \hat{p}, f)$, toujours par orthogonalité de $f - \hat{p}$ et de tout élément de V , donc

$$\|f - Pf\|^2 = (f - Pf, f) = \frac{1}{\det \mathbf{G}_{n+1}} \begin{vmatrix} (\mathbf{b}_0, \mathbf{b}_0) & \cdots & (\mathbf{b}_n, \mathbf{b}_0) & (f, \mathbf{b}_0) \\ \cdots & \cdots & \cdots & \cdots \\ (\mathbf{b}_0, \mathbf{b}_n) & \cdots & (\mathbf{b}_n, \mathbf{b}_n) & (f, \mathbf{b}_n) \\ (\mathbf{b}_0, f) & \cdots & (\mathbf{b}_n, f) & (f, f) \end{vmatrix}.$$

Cette formule donne le carré de la hauteur abaissée de l'extrémité de f sur V .

Soit $\{b_0, \dots, b_n\}$ une base quelconque de V et considérons le paralléloèpe $\mathbf{P}_n = \{x = \theta_0 b_0 + \dots + \theta_n b_n\}$, $0 \leq \theta_0, \dots, \theta_n \leq 1$. Le $(n+1)$ -volume de \mathbf{P}_n est le n -volume de \mathbf{P}_{n-1} multiplié par la hauteur abaissée de l'extrémité de b_n sur le sous-espace sous-tendu par $\{b_0, \dots, b_{n-1}\}$. On adapte la formule précédente:

$$\left(\frac{\text{vol. } \mathbf{P}_n}{\text{vol. } \mathbf{P}_{n-1}} \right)^2 = \frac{\det \mathbf{G}_{n+1}}{\det \mathbf{G}_n},$$

d'où $(n+1)$ -vol. $\mathbf{P}_n = \sqrt{\det \mathbf{G}_{n+1}}$, puisqu'on a bien longueur de $b_0 = [(b_0, b_0)]^{1/2}$ en $n=0$.

Exemple. Monômes dans L^2 , problème de Müntz. Soit $\mathbf{b}_k = x^{\beta_k}$ et le produit scalaire $(f, g) = \int_0^1 f(x)g(x) dx$. Alors, $(\mathbf{b}_k, \mathbf{b}_m) = 1/(\beta_k + \beta_m + 1)$, et le rapport $\det \mathbf{G}_{n+1}/\det \mathbf{G}_n$ est une fonction rationnelle de degré $2n+1$ en β_n , de résidu unité en $\beta_n = -1/2$, positive, et nulle si β_n vaut un des β_k précédents. Cela donne

$$\frac{(\beta_n - \beta_0)^2 \cdots (\beta_n - \beta_{n-1})^2}{(\beta_n + 1/2)(\beta_n + \beta_0 + 1)^2 \cdots (\beta_n + \beta_{n-1} + 1)^2}.$$

Pour l'erreur de meilleure approximation de $f(x) = x^\alpha$ par une combinaison de $x^{\beta_0}, \dots, x^{\beta_n}$, on a

$$\|f - \hat{p}\|^2 = \frac{(\alpha - \beta_0)^2 \cdots (\alpha - \beta_n)^2}{(\alpha + 1/2)(\alpha + \beta_0 + 1)^2 \cdots (\alpha + \beta_n + 1)^2}.$$

Cette norme tend vers 0 quand $n \rightarrow \infty$ si la série des $1/\beta_n$ est divergente (Müntz).

2.5. Factorisation de Cholesky. ⁵

On peut montrer de façon encore plus nette le contenu matriciel de ces manipulations. Constatons que (59) représente en fait l'expression des éléments de l'ancienne base dans la

⁵15 oct. 1875– 31 août 1918, commandant, directeur technique du service géographique de l'armée française. "Officier travailleur, ingénieux, réfléchi, ... , *devra toutefois se méfier comme chef de service de quelque tendance à l'originalité et au paradoxe.* ..." (L^t Col. Hergault, chef d'E.M. de la 7^{ème} Armée, 24 oct. 1916). Cf. C. Brezinski: *André Louis Cholesky*, rapport ANO 347, U.S.T.Lille I, 1995. Aussi dans *Bull. Belg. Math. Soc.- Simon Stevin*, suppl. au n° de déc. 1996, pp. 45-50.

nouvelle base (ce qui est très bien, c'est cette base qui sera utile dans les applications!):

$$[\mathbf{b}_0 \ \mathbf{b}_1 \ \cdots \ \mathbf{b}_n] = [\mathbf{b}_0^\perp \ \mathbf{b}_1^\perp \ \cdots \ \mathbf{b}_n^\perp] \begin{bmatrix} 1 & \alpha_0^{(1)} & \alpha_0^{(2)} & \cdots & \alpha_0^{(n)} \\ & 1 & \alpha_1^{(2)} & \cdots & \alpha_1^{(n)} \\ & & \ddots & & \\ & & & 1 & \alpha_{n-1}^{(n)} \\ & & & & 1 \end{bmatrix}$$

$$\mathbf{b} = \mathbf{b}^\perp \mathbf{M}.$$

Considérons la matrice des produits scalaires (b_i, b_j) (**matrice de Gram**):

$$\mathbf{G}_{n+1} = \begin{bmatrix} (\mathbf{b}_0, \mathbf{b}_0) & \cdots & (\mathbf{b}_0, \mathbf{b}_n) \\ \cdots & \cdots & \cdots \\ (\mathbf{b}_n, \mathbf{b}_0) & \cdots & (\mathbf{b}_n, \mathbf{b}_n) \end{bmatrix} = \left(\begin{bmatrix} \mathbf{b}_0 \\ \vdots \\ \mathbf{b}_n \end{bmatrix}, [\mathbf{b}_0 \ \cdots \ \mathbf{b}_n] \right) = \mathbf{M}^T \left(\begin{bmatrix} \mathbf{b}_0^\perp \\ \vdots \\ \mathbf{b}_n^\perp \end{bmatrix}, [\mathbf{b}_0^\perp \ \cdots \ \mathbf{b}_n^\perp] \right) \overline{\mathbf{M}},$$

donc,

$$\mathbf{G}_{n+1} = \mathbf{M}^T \mathbf{B} \overline{\mathbf{M}},$$

où \mathbf{B} est la matrice diagonale des $\|\mathbf{b}_i^\perp\|^2$.

Soit \mathbf{A} une matrice hermitienne définie positive. On appelle **factorisation de Cholesky** la formation des facteurs du produit

$$\mathbf{A} = \mathbf{L}\mathbf{L}^H$$

où \mathbf{L} est triangulaire inférieure, et où \mathbf{L}^H est la transposée conjuguée de \mathbf{L} . Cette factorisation est unique si on précise les signes des éléments diagonaux de \mathbf{L} . Cette factorisation peut être réalisée par des algorithmes très efficaces (de la même famille que les algorithmes de factorisation LDU liés à l'élimination de Gauss).

Ici, on a donc $\mathbf{A} = \mathbf{G}_{n+1}$ et $\mathbf{L} = \mathbf{M}^T \mathbf{B}^{1/2}$. Si on choisit d'utiliser une base orthonormale, $\mathbf{B} = \mathbf{I}$ et on a donc $\mathbf{M} = \mathbf{L}^T$ fournie par la factorisation; si on choisit de prendre les éléments diagonaux de \mathbf{M} tous égaux à 1, $\mathbf{M} = \mathbf{L}^T \mathbf{B}^{-1/2}$, où \mathbf{B} est la matrice des éléments diagonaux de \mathbf{L} .

Si l'on veut absolument récupérer l'expression de la nouvelle base dans l'ancienne, il faut résoudre $\mathbf{b}^\perp \mathbf{M} = \mathbf{b}$, mais les applications vraiment utiles demandent bien \mathbf{M} plutôt que \mathbf{M}^{-1} : ainsi, s'il faut exprimer un vecteur $v = \sum_0^n c_i \mathbf{b}_i$ dans la nouvelle base,

$$v = \mathbf{b}[c_0, \dots, c_n]^T = \mathbf{b}^\perp \mathbf{M}[c_0, \dots, c_n]^T,$$

le nouveau vecteur de coefficients est bien $\mathbf{M}[c_0, \dots, c_n]^T$.

Volume: puisque hauteur = $\|\mathbf{b}_n^\perp\|$,

$$\text{Vol}_n = \|\mathbf{b}_1^\perp\| \cdots \|\mathbf{b}_n^\perp\| = \text{produit des éléments diagonaux de } L = [\det G_n]^{1/2}.$$

Hadamard: le déterminant de toute matrice $[a_{i,j}]$ est borné par le produit des normes des lignes:

$$|\det A| \leq \prod_{i=1}^n \sqrt{\sum_j |a_{i,j}|^2}.$$

En effet, soit b_i le vecteur de \mathbb{C}^n de composantes $a_{i,1}, \dots, a_{i,n}$ dans la base canonique e_1, \dots, e_n .

Par Gram-Schmidt, on construit la base orthogonale $b_1^\perp, \dots, b_n^\perp$, donc

$$A = \left(\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}, [e_1 \ \cdots \ e_n] \right) = \mathbf{M}^T \left(\begin{bmatrix} b_1^\perp \\ \vdots \\ b_n^\perp \end{bmatrix}, [e_1 \ \cdots \ e_n] \right) = \mathbf{M}^T \begin{bmatrix} \|b_1^\perp\| & & \\ & \ddots & \\ & & \|b_n^\perp\| \end{bmatrix} \mathbf{U},$$

où \mathbf{U} est une matrice unitaire (rotation de base orthonormale), donc de déterminant unité, et $\det A =$ le produit des longueurs des $b_i^\perp \leq$ le produit des longueurs des b_i .

Le déterminant d'une matrice hermitienne définie positive est inférieur au produit des éléments diagonaux.

Exemple: base orthogonale du simplexe $\Sigma_n \subset \mathbb{R}_{n+1}$: lieu des $[t_0, \dots, t_n]$ avec $t_i \geq 0$ et $t_0 + \dots + t_n = 1$. Base de n vecteurs $b_i = [0, 0, \dots, 0, 1, -1, 0, 0, \dots], i = 1, \dots, n$.
Mat de Gram = tridiag(-1,2,-1) = produit de Cholesky de bidiag($-\sqrt{(i-1)/i}, \sqrt{(i+1)/i}$), $i = 1, \dots, n$
un point de $\Sigma_n = (t_0, \dots, t_n) = t_0 b_1 + (t_0 + t_1) b_2 + \dots + (t_0 + \dots + t_{n-1}) b_n + (0, 0, \dots, 1) = u_0 b_1 + \dots + u_{n-1} b_n + (0, 0, \dots, 1)$, avec $0 \leq u_0 \leq u_1 \leq \dots \leq u_{n-1} \leq 1$
 $= u_0 \sqrt{2} b_1^\perp + u_1 (-b_1^\perp / \sqrt{2} + \sqrt{3/2} b_2^\perp) + \dots + u_{n-1} (-\sqrt{(n-1)/n} b_{n-1}^\perp + \sqrt{(n+1)/n} b_n^\perp) + (0, \dots, 0, 1) = (u_0 \sqrt{2} - u_1 / \sqrt{2}) b_1^\perp + \dots + (u_{n-2} \sqrt{n/(n-1)} - u_{n-1} \sqrt{(n-1)/n}) b_{n-1}^\perp + u_{n-1} \sqrt{(n+1)/n} b_n^\perp + (0, 0, \dots, 0, 1)$

$$dS = \sqrt{n+1} du_0 \dots du_{n-1}, \quad \text{volume} = \sqrt{n+1}/n!$$

cf. <http://en.wikipedia.org/wiki/Simplex>

3. Polynômes orthogonaux.

On applique maintenant ce qui précède au cas où X est un espace préhilbertien de fonctions, muni d'un produit scalaire du type (50) ou, plus généralement (53) : $(f, g) = \int_a^b f(x)g(x) d\mu(x)$, et où $V = \mathcal{P}_n$.

3.1. Construction d'une base orthogonale de \mathcal{P}_n .

L'espace vectoriel de dimension finie \mathcal{P}_n (de dimension $n+1$) contient une infinité de bases orthogonales. Ici, on se donne la contrainte supplémentaire: degré $\Phi_i = i$, $i = 0, 1, \dots, n$, donc, en partant de x^0, x, \dots, x^n :

$$\begin{aligned} \Phi_0(x) &= 1 \\ \Phi_1(x) &= x + \Phi_1(0) \\ \Phi_2(x) &= x^2 + A'_2 x + \Phi_2(0) \\ &\dots \quad \dots \quad \dots \quad \dots \\ \Phi_n(x) &= x^n + A'_n x^{n-1} + \dots + \Phi_n(0) \end{aligned}$$

3.1.1. *Moments, matrice de Hankel.* Exprimons que

$$\Phi_i(x) = \Phi_i(0) + \dots + A'_i x^{i-1} + x^i$$

est orthogonal à $1, x, x^2, \dots, x^{i-1}$. On retrouve une matrice de Gram:

$$\begin{array}{cccccc} \Phi_i(0)(x^0, x^0) & + & \cdots & + & A'_i(x^{i-1}, x^0) & + & (x^i, x^0) & = & 0 \\ \Phi_i(0)(x^0, x^1) & + & \cdots & + & A'_i(x^{i-1}, x^1) & + & (x^i, x^1) & = & 0 \\ \cdots & & \cdots & & \cdots & & \cdots & & \\ \Phi_i(0)(x^0, x^{i-1}) & + & \cdots & + & A'_i(x^{i-1}, x^{i-1}) & + & (x^i, x^{i-1}) & = & 0 \end{array}$$

Soit $(x^i, x^j) = \int_a^b x^{i+j} d\mu(x) =: \mu_{i+j}$, **moment**⁶ d'ordre $i+j$ (autour de 0) de $d\mu$. Ces moments existent tant que $i+j \leq 2n$, par hypothèse et la discussion "Existence de l'intégrale" faite au point 1, p. 74. On a donc

$$\begin{bmatrix} \mu_0 & \mu_1 & \cdots & \mu_{i-1} & \mu_i \\ \mu_1 & \mu_2 & \cdots & \mu_i & \mu_{i+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mu_{i-1} & \mu_i & \cdots & \mu_{2i-2} & \mu_{2i-1} \end{bmatrix} \begin{bmatrix} \Phi_i(0) \\ \Phi'_i(0) \\ \vdots \\ A'_i \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (60)$$

système de i équations à i inconnues $\Phi_i(0), \Phi'_i(0), \dots, A'_i$. La matrice carrée \mathbf{H}

$$\mathbf{H}_i = \begin{bmatrix} \mu_0 & \mu_1 & \cdots & \mu_{i-1} \\ \mu_1 & \mu_2 & \cdots & \mu_i \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{i-1} & \mu_i & \cdots & \mu_{2i-2} \end{bmatrix} \quad (61)$$

est *définie positive*: soit $\mathbf{v} = [v_0, v_1, \dots, v_{i-1}]^T$, avec au moins un des $v_j \neq 0$, alors,

$$\mathbf{v}^T \mathbf{H} \mathbf{v} = \sum_{p=0}^{i-1} v_p \sum_{q=0}^{i-1} \mu_{p+q} v_q = \sum_{p=0}^{i-1} v_p \int_a^b \left(\sum_{q=0}^{i-1} v_q x^{p+q} \right) d\mu(x) = \int_a^b \left(\sum_{p=0}^{i-1} v_p x^p \right)^2 d\mu(x) > 0$$

car le produit scalaire est défini positif. On aura d'ailleurs reconnu la **matrice de Gram** de la base $\{\mathbf{b}_j = x^j\}_0^{i-1}$ de \mathcal{P}_{i-1} (cf. (54) p. 77).

L'ensemble des polynômes de degré i orthogonaux à \mathcal{P}_{i-1} est donné par

$$\{A_i \Phi_i\}, \quad A_i \neq 0, (\Phi_i(x) = x^i + \cdots).$$

Les polynômes *orthonormaux* $\|\varphi_i\| = 1$ sont donnés par $A_i \Phi_i$ tels que $\|A_i \Phi_i\| = |A_i| \|\Phi_i\| = 1$:

$$\varphi_i = A_i \Phi_i = \pm \frac{\Phi_i}{\|\Phi_i\|}.$$

Exemples. Des calculs simples, mais devenant vite un peu fastidieux, aboutissent à l'explicitation des Φ_i pour de petites valeurs de i :

On voit apparaître d'intéressants déterminants...

⁶La notion de moment se retrouve également en *mécanique* et en *probabilités*. Le moment d'ordre k autour de x_0 de $d\mu$ est $\int_a^b (x - x_0)^k d\mu(x)$.

degré	$\Phi_i(x)$	$\varphi_i(x)$
0	1	$\pm \frac{1}{\sqrt{\mu_0}}$
1	$x - \frac{\mu_1}{\mu_0}$	$\pm \frac{x - \frac{\mu_1}{\mu_0}}{\sqrt{\mu_2 - \frac{\mu_1^2}{\mu_0}}}$
2	$x^2 - \frac{\mu_0\mu_3 - \mu_1\mu_2}{\mu_0\mu_2 - \mu_1^2}x + \frac{\mu_1\mu_3 - \mu_2^2}{\mu_0\mu_2 - \mu_1^2}$	$\pm \frac{x^2 - \frac{\mu_0\mu_3 - \mu_1\mu_2}{\mu_0\mu_2 - \mu_1^2}x + \frac{\mu_1\mu_3 - \mu_2^2}{\mu_0\mu_2 - \mu_1^2}}{\sqrt{\frac{\mu_0\mu_2\mu_4 - \mu_1^2\mu_4 - \mu_0\mu_3^2 + 2\mu_1\mu_2\mu_3 - \mu_2^3}{\mu_0\mu_2 - \mu_1^2}}}$

TABLE 1. Polynômes orthogonaux de petits degrés.

3.1.2. Déterminants. Ajoutons

$$[1 \quad x \quad \dots \quad x^{i-1} \quad x^i] \begin{bmatrix} \Phi_i(0) \\ \Phi_i'(0) \\ \vdots \\ A_i' \\ 1 \end{bmatrix} = \Phi_i(x)$$

à (60) pour obtenir un système de $i + 1$ équations à $i + 1$ inconnues, et extrayons “l’inconnue” 1 par formule de Cramer:

$$1 = \frac{\begin{vmatrix} \mu_0 & \mu_1 & \dots & \mu_{i-1} & 0 \\ \mu_1 & \mu_2 & \dots & \mu_i & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \mu_{i-1} & \mu_i & \dots & \mu_{2i-2} & 0 \\ 1 & x & \dots & x^{i-1} & \Phi_i(x) \end{vmatrix}}{\begin{vmatrix} \mu_0 & \mu_1 & \dots & \mu_{i-1} & \mu_i \\ \mu_1 & \mu_2 & \dots & \mu_i & \mu_{i+1} \\ \dots & \dots & \dots & \dots & \dots \\ \mu_{i-1} & \mu_i & \dots & \mu_{2i-2} & \mu_{2i-1} \\ 1 & x & \dots & x^{i-1} & x^i \end{vmatrix}}$$

soit:

$$\Phi_i(x) = \frac{\begin{vmatrix} \mu_0 & \mu_1 & \dots & \mu_{i-1} & \mu_i \\ \mu_1 & \mu_2 & \dots & \mu_i & \mu_{i+1} \\ \dots & \dots & \dots & \dots & \dots \\ \mu_{i-1} & \mu_i & \dots & \mu_{2i-2} & \mu_{2i-1} \\ 1 & x & \dots & x^{i-1} & x^i \end{vmatrix}}{\det \mathbf{H}_i}.$$

Le numérateur est linéaire en les éléments de la dernière ligne, donc, $\forall f$,

$$(\Phi_i, f) = \frac{\begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_{i-1} & \mu_i \\ \mu_1 & \mu_2 & \cdots & \mu_i & \mu_{i+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mu_{i-1} & \mu_i & \cdots & \mu_{2i-2} & \mu_{2i-1} \\ (1, f) & (x, f) & \cdots & (x^{i-1}, f) & (x^i, f) \end{vmatrix}}{\det \mathbf{H}_i}.$$

En particulier, $(\Phi_i, \Phi_i) = (\Phi_i, x^i + g)$, avec $g \in \mathcal{P}_{i-1}$, donc, $(\Phi_i, \Phi_i) = (\Phi_i, x^i)$:

$$(\Phi_i, \Phi_i) = \|\Phi_i\|^2 = \frac{\det \mathbf{H}_{i+1}}{\det \mathbf{H}_i},$$

ce qui confirme d'ailleurs que $\det \mathbf{H}_i > 0$ (la formule est valable à partir de $i = 0$ si on pose $\det \mathbf{H}_0 = 1$).

Remarque. Si μ n'a que n points de croissance x_1, x_2, \dots, x_n sur $[a, b]$, on peut encore déterminer Φ_n mais $\|\Phi_n\| = 0$ (c'est-à-dire $\Phi_n(x) = (x - x_1) \dots (x - x_n)$), car \mathbf{H}_{n+1} n'est que semi-définie positive: il existe $\mathbf{v} = [v_0, \dots, v_n]^T \neq 0$ tel que $\mathbf{H}_{n+1}\mathbf{v} = 0$, les v_j ne sont autres que les coefficients de $(x - x_1) \dots (x - x_n) = \Phi_n(x)$.

3.1.3. *Relation avec la méthode de Gram-Schmidt.* On a appliqué (59) (p. 81) à $V = \mathcal{P}_n$ de base initiale $\mathbf{b}_i = x^i$, $i = 0, 1, \dots, n$. On cherche à construire une base orthogonale $\mathbf{b}_i^\perp = \Phi_i$ qui soit un *réarrangement triangulaire* de la base donnée. On utilise les Φ_i à mesure qu'on les construit:

- 1) $\Phi_0 = 1$;
- 2) pour $i > 0$, supposons disposer de $\Phi_0, \dots, \Phi_{i-1}$, alors Φ_i doit être une combinaison linéaire

$$\Phi_i = \alpha_i^{(i)} \Phi_0 + \cdots + \alpha_i' \Phi_{i-1} + x^i \quad (62)$$

orthogonale à $\Phi_0, \dots, \Phi_{i-1}$:

$$0 = (\Phi_i, \Phi_j) = (\alpha_i^{(i)} \Phi_0 + \cdots + \alpha_i' \Phi_{i-1} + x^i, \Phi_j) = \alpha_i^{(i-j)} \|\Phi_j\|^2 + (x^i, \Phi_j),$$

donc, le coefficient de Φ_j est

$$\alpha_i^{(i-j)} = -\frac{(x^i, \Phi_j)}{\|\Phi_j\|^2}, \quad j = 0, \dots, i-1.$$

On accède donc progressivement à tous les éléments de la nouvelle base.

On verra des algorithmes plus efficaces, basés sur la *relation de récurrence*.

Conclusion de cette partie. La détermination d'une base orthogonale de \mathcal{P}_n n'est qu'un exemple de construction de base orthogonale d'un espace de dimension finie. Cette construction passe par l'orthogonalisation d'une base donnée (méthode de Gram-Schmidt) et revient à considérer la factorisation de Cholesky de la matrice de Gram de la base initiale. Dans le cas des polynômes, et avec un produit scalaire de la forme $(f, g) = \int_a^b f(x)g(x)d\mu(x)$, si la base initiale est celle des monômes de \mathcal{P}_n , la matrice de Gram est constituée des moments μ_{i+j} (**matrice de Hankel**).

3.2. Relation de récurrence.

Le résultat suivant est de la plus haute importance pour tout ce qui concerne l'utilisation des polynômes orthogonaux:

Théorème. *Toute suite de polynômes orthogonaux $\{\Phi_0, \Phi_1, \dots\}$, $\Phi_i(x) = x^i + \dots$, relatifs à un produit scalaire de la forme $(f, g) = \int_a^b f(x)g(x) d\mu(x)$ vérifie une relation de récurrence d'ordre deux:*

$$\Phi_{n+1}(x) = (x - \alpha_n)\Phi_n(x) - \beta_n^2 \Phi_{n-1}(x), \quad n = 1, 2, \dots \quad (63)$$

avec $\Phi_0 = 1$ et $\Phi_1(x) = x - \alpha_0$,

$$\alpha_n = \frac{(x \Phi_n, \Phi_n)}{\|\Phi_n\|^2} = \frac{\int_a^b x \Phi_n^2(x) d\mu(x)}{\int_a^b \Phi_n^2(x) d\mu(x)}, \quad n = 0, 1, \dots$$

$$\beta_n^2 = \frac{\|\Phi_n\|^2}{\|\Phi_{n-1}\|^2} = \frac{\int_a^b \Phi_n^2(x) d\mu(x)}{\int_a^b \Phi_{n-1}^2(x) d\mu(x)}, \quad n = 1, 2, \dots \quad (64)$$

Si μ n'a que $N < \infty$ points de croissance sur $[a, b]$, les relations de récurrence sont limitées à $1 \leq n \leq N - 1$.

En effet, effectuons la division du polynôme Φ_{n+1} de degré $n + 1$ par Φ_n de degré n , soit $x - \alpha_n$ le quotient, et R_n le reste qui doit être de degré $< n$: soit $R_n = \gamma_n \Phi_{n-1} + \gamma'_n \Phi_{n-2} + \dots$,

$$\Phi_{n+1}(x) = (x - \alpha_n)\Phi_n(x) + \gamma_n \Phi_{n-1}(x) + \gamma'_n \Phi_{n-2}(x) + \dots,$$

et formons un produit scalaire en multipliant à gauche par

- (1) Φ_n : $0 = (\Phi_n(x), x\Phi_n(x)) - \alpha_n(\Phi_n(x), \Phi_n(x))$, ce qui donne α_n .
- (2) Φ_{n-1} : $0 = (\Phi_{n-1}(x), x\Phi_n(x)) + \gamma_n(\Phi_{n-1}(x), \Phi_{n-1}(x))$, que l'on transforme encore en utilisant $(\Phi_{n-1}(x), x\Phi_n(x)) = (x\Phi_{n-1}(x), \Phi_n(x))$, et $x\Phi_{n-1}(x) = \Phi_n(x) + \psi(x)$ avec $\psi \in \mathcal{P}_{n-1}$, donc orthogonal à Φ_n :

$$\gamma_n = -\|\Phi_n\|^2 / \|\Phi_{n-1}\|^2$$

noté $-\beta_n^2$ dans (63).

- (3) Φ_{n-j} avec $j > 1$: $0 = (\Phi_{n-j}(x), x\Phi_n(x)) + \gamma_n^{(j)}(\Phi_{n-j}(x), \Phi_{n-j}(x))$, or $(\Phi_{n-j}(x), x\Phi_n(x)) = (x\Phi_{n-j}(x), \Phi_n(x))$ et $x\Phi_{n-j}(x)$ n'est que de degré $n - j + 1 < n$, donc orthogonal à Φ_n , R_n n'a donc pas d'autre composante non nulle que celle de Φ_{n-1} . \square

Remarque. On a utilisé plusieurs fois dans la preuve la propriété $(f(x), xg(x)) = (xf(x), g(x)) = \int_a^b xf(x)g(x)d\mu(x)$. On a ainsi mis en évidence le caractère *formellement autoadjoint* de l'opérateur de multiplication par x pour ce type de produit scalaire. On reviendra sur cette notion.

Si on choisit la suite tout aussi orthogonale $\{A_n \Phi_n\}$, avec des constantes $A_n \neq 0$, la récurrence devient

$$A_{n+1}\Phi_{n+1}(x) = \frac{A_{n+1}}{A_n}(x - \alpha_n)A_n\Phi_n(x) - \frac{A_{n+1}}{A_{n-1}}\beta_n^2 A_{n-1}\Phi_{n-1}(x),$$

Pour les *polynômes orthonormaux*, $A_n = 1/\|\Phi_n\|$, de sorte que $A_n/A_{n+1} = \|\Phi_{n+1}\|/\|\Phi_n\| = \beta_{n+1}$, par (64), et si on choisit de prendre $\beta_n > 0$,

$$\beta_{n+1}\varphi_{n+1}(x) = (x - \alpha_n)\varphi_n(x) - \beta_n\varphi_{n-1}(x). \quad (65)$$

3.3. Quelques algorithmes utilisant la récurrence.

Voici, dans une syntaxe vaguement inspirée de `matlab`, quelques échantillons de la souplesse et de la puissance des relations de récurrence. On suppose disposer des valeurs numériques de α_i, β_i pour $i = 0, 1, \dots, n$:

- (1) Accéder à la valeur numérique de $\Phi_n(x)$:

```
p = 1; q = x - alpha_0;
for i = 1 : 1 : n - 1;
r = (x - alpha_i)q - beta_i^2*p; %ici, p = Phi_{i-1}(x), q = Phi_i(x), r = Phi_{i+1}(x)
p = q; q = r;
end;
La réponse est r.
```

- (2) Effectuer la somme $s = \sum_{i=0}^n c_i \Phi_i(x)$:

- (a) Forme progressive: on ajoute l'accumulation des sommes partielles à l'algorithme précédent:

```
p = 1; q = x - alpha_0; s = c_0 + c_1*q;
for i = 1 : 1 : n - 1;
r = (x - alpha_i)q - beta_i^2*p; %ici, p = Phi_{i-1}(x), q = Phi_i(x), r = Phi_{i+1}(x)
s = s + c_{i+1}*r;
p = q; q = r;
end;
```

- (b) Forme régressive: on écrit (63) jusque Φ_n , suivi de l'expression de s :

$$\mathbf{R}\boldsymbol{\varphi} = \mathbf{se} : \begin{bmatrix} x - \alpha_0 & -1 & & & & \\ -\beta_1^2 & x - \alpha_1 & -1 & & & \\ & \dots & \dots & \dots & & \\ & & & -\beta_{n-1}^2 & x - \alpha_{n-1} & -1 \\ c_0 & c_1 & \dots & c_{n-1} & c_n & \end{bmatrix} \begin{bmatrix} \Phi_0 \\ \Phi_1(x) \\ \vdots \\ \Phi_{n-1}(x) \\ \Phi_n(x) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ s \end{bmatrix}$$

et on multiplie par le vecteur ligne $\boldsymbol{\sigma} = [s_0, s_1, \dots, s_n]$ avec $s_n = 1$, et toutes les composantes de $\boldsymbol{\sigma}\mathbf{R}$ égales à 0, sauf la première, alors, $s = (x - \alpha_0)s_0 - \beta_1^2 s_1 + c_0$:

```
q = 0; r = 0;
for i = n : -1 : 0;
p = (x - alpha_i)q - beta_{i+1}^2*r + c_i; % ici, p = s_{i-1}, q = s_i, r = s_{i+1}
r = q; q = p;
end;
s = p;
```

- (3) Accéder à la valeur numérique de la dérivée $\Phi'_n(x)$: il suffit de dériver (63) pour construire la récurrence (non homogène) des dérivées $\Phi'_{n+1}(x) = (x - \alpha_n)\Phi'_n(x) - \beta_n^2\Phi'_{n-1}(x) + \Phi_n(x)$

```
p = 1; q = x - alpha_0;
p1 = 0; q1 = 1;
```

```

for i = 1 : 1 : n - 1;
r = (x - alpha_i)q - beta_i^2 p; %ici, p = Phi_{i-1}(x), q = Phi_i(x), r = Phi_{i+1}(x)
r1 = (x - alpha_i)q1 - beta_i^2 p1 + q; %ici, p1 = Phi'_{i-1}(x), q1 = Phi'_i(x), r1 = Phi'_{i+1}(x)
p = q; q = r;
p1 = q1; q1 = r1;
end;

```

La réponse est r1.

(4) Réarranger $\sum_{i=0}^n c_i \Phi_i(x) = \sum_{i=0}^n a_i x^i$.

(a) Passer des a_i aux c_i : on constitue progressivement la somme par les différentes étapes du schéma de Horner $s = xs + a_i$, $i = n, n-1, \dots$ (Hamming) en utilisant la récurrence (63) sous la forme $x \Phi_i(x) = \Phi_{i+1}(x) + \alpha_i \Phi_i(x) + \beta_i^2 \Phi_{i-1}(x)$:

```

c_{n+1} = 0; c_{n+2} = 0;
for i = n : -1 : 0;
% ici, c_{i+1} \Phi_0 + \dots + c_n \Phi_{n-i-1} = a_{i+1} + \dots + a_n x^{n-i-1}
c_i = a_i + \alpha_0 c_{i+1} + \beta_1^2 c_{i+2};
for j = 1 : 1 : n - i;
c_{i+j} = c_{i+j} + \alpha_j c_{i+j+1} + \beta_{j+1}^2 c_{i+j+2};
end;
end;
for i = 0 : 1 : n; a_i = a_i / A_i; end;

```

(b) Passer des c_i aux a_i : on reprend les s_i du point 2b ci-dessus, en constatant que s_i est un polynôme de degré $n-1-i$, $i = n-1, n-2, \dots, -1$:

```

% polynome dans la base x^n x^{(n-1)} ... 1
r=zeros(1,n+1);r(n+1)=c(n+1);
if n>0
q=zeros(1,n+1);q(n)=r(n+1);
q(n+1)=-alpha(n)*r(n+1)+c(n);p=zeros(1,n+1);
for i=n-1:-1:1
p(i:n)=q(i+1:n+1);p(n+1)=0; % coefficients de x s_{i-1} (shift)
p=p-alpha(i)*q-beta2(i+1)*r;p(n+1)=p(n+1)+c(i); % p contient maintenant
les coefficients de s_{i-2} (alpha(i) vaut alpha_{i-1} etc.)
r=q;q=p;
end;
r=q;
end;
a=r;

```

On voit l'importance de la récurrence dans tout traitement numérique lié aux polynômes orthogonaux. Encore faut-il connaître les coefficients de la récurrence...

À partir de la fonction de poids (ou densité) w , ou de la fonction croissante μ , on peut construire⁷ les coefficients α_n et β_n de (63) soit par

- des méthodes de type “Stieltjes”, où on évalue α_n et β_n par (64) à partir de valeurs numériques de Φ_n (elles-mêmes obtenues par exploitation de la récurrence jusqu’au

⁷W. Gautschi, On generating orthogonal polynomials, *SIAM J. Sci. Stat. Comput.* **3** (1982) 289-317.

degré n) en utilisant une formule d'intégration numérique (ou une évaluation exacte si μ est une fonction en escalier),

- des méthodes de type “Tchebycheff”, où on manipule des intégrales définies (en partant des moments $\mu_k = \int_a^b x^k d\mu(x)$, ou, mieux encore, de moments modifiés⁸ $\int_a^b p_k(x) d\mu(x)$ avec des polynômes p_k “bien choisis”.

Voici l'algorithme le plus primitif de cette famille, on construit les intégrales $\mu_{k,m} := \int_a^b x^k \Phi_m(x) d\mu(x)$ en partant de $\mu_{k,0} = \mu_k, k = 0, 1, \dots, 2n, \mu_{k,1} = \int_a^b x^k (x - \alpha_0) d\mu(x) = \mu_{k+1} - \alpha_0 \mu_k, k = 0, 1, \dots, 2n - 1$, où $\alpha_0 = \mu_1/\mu_0$. Ensuite:

$$\begin{aligned} \mu_{k,m+1} &= \int_a^b x^k \Phi_{m+1}(x) d\mu(x) \\ &= \int_a^b x^k [(x - \alpha_m) \Phi_m(x) - \beta_m^2 \Phi_{m-1}(x)] d\mu(x), \quad (\text{par (63)}) \\ &= \mu_{k+1,m} - \alpha_m \mu_{k,m} - \beta_m^2 \mu_{k,m-1} \end{aligned}$$

pour $k = 0, 1, \dots, 2n - m - 1$, et on s'arrête quand $m = n$. Il faut connaître α_m et β_m ! L'astuce est de remarquer que $\mu_{k,m} = 0$ si $m > k$ par orthogonalité. On commence en fait en $k = m - 1$, où on a: $\mu_{m,m} - \beta_m^2 \mu_{m-1,m-1}$, ce qui donne β_m . Ensuite, en $k = m, 0 = \mu_{m+1,m} - \alpha_m \mu_{m,m} - \beta_m^2 \mu_{m-1,m-1}$, d'où α_m .

Programmes matlab: voir MATLAB SUITE FOR GENERATING ORTHOGONAL POLYNOMIALS <http://www.cs.purdue.edu/archives/2002/wxg/codes/OPQ.html>

On aurait pu obtenir ces informations par les algorithmes généraux d'orthogonalisation (Gram-Schmidt, Cholesky), mais la factorisation de Cholesky d'une matrice d'ordre n coûte $\approx n^3/6$ opérations, alors que ce que l'on vient de voir ne représente que $\approx n^2$ opérations. Ces simplifications, dont on ne comprend pas très bien l'origine⁹ semblent dues à une association favorable du caractère polynomial des fonctions de base et du type de produit scalaire utilisé.

Exemple: polynômes de Legendre sur $[a, b]$. Prenons $w(x) = 1$ (donc $\mu(x) = x +$ constante) sur un intervalle borné $[a, b]$. On a $\mu_{k,0} = \mu_k = \int_a^b x^k dx = (b^{k+1} - a^{k+1})/(k+1)$, $\alpha_0 = \mu_1/\mu_0 = (a+b)/2$, $\mu_{k,1} = \mu_{k+1} - \alpha_0 \mu_k = [k(b^{k+2} - a^{k+2}) - (k+2)ab(b^k - a^k)]/(2(k+1)(k+2))$. Après quelques calculs pénibles (on aurait mieux fait de prendre $p_k(x) = (x-a)^k$ ou $p_k(x) = (x - (a+b)/2)^k$), on trouve $\alpha_1 = \alpha_2 = \dots = (a+b)/2$, $\beta_1^2 = (b-a)^2/12$, $\beta_2^2 = (b-a)^2/15$, on verra plus loin la formule donnant tous les β_n dans ce cas...

Algorithme quotient-différence (qd) de Rutishauser-Stiefel.

Voici un algorithme beaucoup plus subtil reliant les moments (constantes de Schwarz) aux coefficients de la récurrence (constantes de Lanczos). Pour éviter des divisions par zéro, partons des moments par rapport à a : $\mu_k^{(a)} := \int_a^b (x-a)^k d\mu(x)$, $k = 0, 1, \dots$. On construit alors les colonnes successives du tableau

⁸Le problème de la détermination de α_n et β_n (constantes de Lanczos) en fonction des moments $\mu_k = \int_a^b x^k d\mu(x)$ (constantes de Schwarz) est mal conditionné si n est grand. Pour une synthèse récente sur la question, cf. W. Gautschi, Algorithm 726. ORTHPOL: a package of routines for generating orthogonal polynomials and Gauss-type quadrature rules, *ACM Trans. Math. Soft.* **20** (1994) 21-62.

⁹Pour les polynômes orthogonaux relativement à un produit scalaire de Sobolev $(f, g) = \int_a^b [f(x)g(x)d\mu_1(x) + f'(x)g'(x)d\mu_2(x)]$, on ne connaît rien de tel...

$$\begin{array}{ccccccc}
 & & & & & & \text{par } q_1^{(k)} := \mu_{k+1}^{(a)}/\mu_k^{(a)}, k = 0, 1, \dots \text{ et les } \mathbf{r\grave{e}gles} \\
 & & & & & & \mathbf{du losange} \\
 e_0^{(1)} = 0 & q_1^{(0)} & & & & & e_m^{(n)} := q_m^{(n+1)} - q_m^{(n)} + e_{m-1}^{(n+1)}, \\
 & e_1^{(0)} & & & & & \\
 e_0^{(2)} = 0 & q_1^{(1)} & q_2^{(0)} & & & & q_{m+1}^{(n)} := \frac{e_m^{(n+1)}}{e_m^{(n)}} q_m^{(n+1)}, \\
 & e_1^{(1)} & e_2^{(0)} & & & & \\
 e_0^{(3)} = 0 & q_1^{(2)} & q_2^{(1)} & \cdots & & & \text{pour } m = 1, 2, \dots, n = 0, 1, \dots \\
 & e_1^{(2)} & e_2^{(1)} & & & & \\
 e_0^{(4)} = 0 & q_1^{(3)} & q_2^{(2)} & \cdots & & & \\
 & e_1^{(3)} & e_2^{(2)} & & & & \\
 \vdots & \vdots & \vdots & \cdots & & & \\
 \vdots & \vdots & \vdots & \cdots & & &
 \end{array} \tag{66}$$

Alors, $\alpha_k = a + q_{k+1}^{(0)} + e_k^{(0)}$, $k = 0, 1, \dots$, $\beta_k^2 = e_k^{(0)} q_k^{(0)}$, $k = 1, 2, \dots$

Explication¹⁰: soit $\Phi_{k,n}$ le polynôme orthogonal monique de degré k par rapport à $(x-a)^n d\mu(x)$ sur (a, b) (*polynômes de Hadamard*). $\Phi_k^{(n+1)} - \Phi_k^{(n)}$ est 1) de degré $\leq k-1$; 2) orthogonal à \mathcal{P}_{k-2} par rapport à $(x-a)^{n+1} d\mu(x)$, donc $\Phi_k^{(n+1)} - \Phi_k^{(n)} =$ une constante fois $\Phi_{k-1}^{(n+1)}$, soit $\Phi_k^{(n+1)} - \Phi_k^{(n)} = -e_k^{(n)} \Phi_{k-1}^{(n+1)}$. $\Phi_{k+1}^{(n)}(x) - (x-a)\Phi_k^{(n+1)}$ est 1) de degré $\leq k$; 2) orthogonal à \mathcal{P}_{k-1} par rapport à $(x-a)^n d\mu(x)$, donc $\Phi_{k+1}^{(n)}(x) - (x-a)\Phi_k^{(n+1)} =$ une constante fois $\Phi_k^{(n)}$, soit $\Phi_{k+1}^{(n)}(x) - (x-a)\Phi_k^{(n+1)} = -q_{k+1}^{(n)} \Phi_k^{(n)}$. Remarquons que $q_k^{(n)} = -\Phi_k^{(n)}(a)/\Phi_{k-1}^{(n)}(a)$.

On retrouve la forme de la récurrence des $\Phi_k^{(n)}$:

$$\begin{aligned}
 0 &= \Phi_{k+1}^{(n)} - (x-a)\Phi_k^{(n+1)} + q_{k+1}^{(n)} \Phi_k^{(n)} + e_k^{(n)} \left[\Phi_k^{(n)} - (x-a)\Phi_{k-1}^{(n+1)} + q_k^{(n)} \Phi_{k-1}^{(n)} \right]; \\
 \Phi_{k+1}^{(n)}(x) &= \left(x-a - q_{k+1}^{(n)} - e_k^{(n)} \right) \Phi_k^{(n)}(x) - e_k^{(n)} q_k^{(n)} \Phi_{k-1}^{(n)}(x).
 \end{aligned}$$

Pour retrouver les formules du losange, formons les produits scalaires $(\Phi_k^{(n)}, \Phi_{k-1}^{(n+1)})_{n+1} = e_k^{(n)} \|\Phi_{k-1}^{(n+1)}\|_{n+1}^2$, ou $e_k^{(n)} = \|\Phi_k^{(n)}\|_n^2 / \|\Phi_{k-1}^{(n+1)}\|_{n+1}^2$; $((x-a)\Phi_k^{(n+1)}, \Phi_k^{(n)})_n = q_{k+1}^{(n)} \|\Phi_k^{(n)}\|_n^2$, ou $q_{k+1}^{(n)} = \|\Phi_k^{(n+1)}\|_{n+1}^2 / \|\Phi_k^{(n)}\|_n^2$. On retrouve d'abord ainsi la deuxième règle (66).

On retrouve les deux règles (66) en formant la récurrence des $\Phi_k^{(n+1)}$

$$\begin{aligned}
 0 &= \Phi_{k+1}^{(n+1)} - \Phi_{k+1}^{(n)} + e_{k+1}^{(n)} \Phi_k^{(n+1)} + q_{k+1}^{(n)} \left[\Phi_k^{(n+1)} - \Phi_k^{(n)} + e_k^{(n)} \Phi_{k-1}^{(n+1)} \right]; \\
 \Phi_{k+1}^{(n+1)}(x) &= (x-a - q_{k+1}^{(n)} - e_{k+1}^{(n)}) \Phi_k^{(n+1)}(x) - e_k^{(n)} q_{k+1}^{(n)} \Phi_{k-1}^{(n+1)}(x).
 \end{aligned}$$

Exercice. Montrez que les polynômes $\{\Psi_{2k}(y) = \Phi_{k,n}(a+y^2), \Psi_{2k+1}(y) = y\Phi_{k,n+1}(a+y^2)\}$ sont orthogonaux sur $-\sqrt{b-a}, \sqrt{b-a}$ par rapport à $y^{2n} d[\text{sign } y (\mu(a+y^2) - \mu(a))]$.

Et on a la récurrence $\Psi_{m+1}(y) = y\Psi_m(y) - \gamma_m^2 \Psi_{m-1}(y)$, avec $\gamma_{2k}^2 = e_k^{(n)}$, $\gamma_{2k+1}^2 = q_{k+1}^{(n)}$.

Exemples: polynômes de Legendre sur $[a, b]$.

Reprenons l'exemple précédent: $\mu_k^{(a)} = \int_a^b (x-a)^k dx = \frac{(b-a)^{k+1}}{k+1}$. $q_1^{(k)} = (b-a)(k+1)/(k+2)$, $e_1^{(k)} = (b-a)/((k+2)(k+3))$, $q_2^{(k)} = (b-a)(k+2)^2/((k+3)(k+4))$, $e_2^{(k)} = 4(b-a)/((k+4)(k+5))$, $q_3^{(k)} = (b-a)(k+3)^2/((k+5)(k+6))$, etc.:
 $q_n^{(k)} = (b-a)(k+n)^2/((k+2n-1)(k+2n))$, $e_n^{(k)} = n^2(b-a)/((k+2n)(k+2n+1))$,
 $\alpha_n = (a+b)/2$, $\beta_n^2 = \frac{n^2(b-a)^2}{4(4n^2-1)}$.

¹⁰Reprise de P. Henrici, *Applied and Computational Complex Analysis*, vol. 1, Wiley, 1974, § 7.6, 7.7.

Polynômes de Laguerre $L_n^{(\alpha)}$.

$a = 0, b = \infty, d\mu(x) = x^\alpha e^{-x}$. $\mu_k = \int_0^\infty x^{k+\alpha} e^{-x} dx = \Gamma(k + \alpha + 1)$. $q_1^{(k)} = \Gamma(k + \alpha + 2)/\Gamma(k + \alpha + 1) = k + \alpha + 1$. On montre facilement $q_n^{(k)} = k + \alpha + n, e_n^{(k)} = n, \alpha_n = \alpha + 2n + 1, \beta_n^2 = n(\alpha + n)$.

Polynômes d’Hermite H_n . Densité e^{-x^2} sur $(-\infty, \infty)$. D’après l’exercice précédent, $\mu(x = y^2) = \int_0^y e^{-t^2} dt, y \geq 0$, donc $d\mu(x) = \frac{1}{2} x^{-1/2} e^{-x}$, $H_{2n}(y) = L_n^{(-1/2)}(y^2), H_{2n+1}(y) = y L_n^{(1/2)}(y^2), \alpha_n = 0, \beta_n^2 = n/2$.

3.4. Zéros des polynômes orthogonaux.

Un polynôme de degré n a au plus n zéros réels. Les choses sont beaucoup plus précises pour les polynômes orthogonaux:

Théorème. Toute fonction continue de norme > 0 , orthogonale à \mathcal{P}_{n-1} selon un produit scalaire $(f, g) = \int_a^b f(x)g(x)d\mu(x)$, où μ a au moins n points de croissance sur $[a, b]$, doit admettre au moins n changements de signe sur l’intervalle ouvert (a, b) .

En particulier, le polynôme orthogonal de degré n a n zéros simples dans (a, b) .

En effet, supposons que f ne change de signe qu’aux points $a < x_1 < \dots < x_k < b$ avec $k < n$. Pour fixer les idées, soit $f(x) \geq 0$ pour $x_k < x < b, f(x) \leq 0$ pour $x_{k-1} < x < x_k$, etc. Alors, on aurait $f(x)p(x) \geq 0$ sur tout $[a, b]$, avec $p(x) = (x - x_1) \dots (x - x_k) \in \mathcal{P}_{n-1}$, donc $\int_a^b f(x)p(x) d\mu(x) = 0$ avec $f(x)p(x) \geq 0$: fp devrait être nulle en chaque point de croissance de $\mu \Rightarrow f$ devrait être nulle en chacun de ces points ($f(x)p(x) = 0 \Rightarrow f(x) = 0$ si $p(x) \neq 0$ et nous savons que $f(x) = 0$ si $p(x) = 0$), et on aurait $\|f\| = 0$. \square

Exercice. Montrez que, si le support de μ est constitué de plusieurs composantes connexes, Φ_n possède au plus un zéro simple dans tout intervalle séparant deux de ces composantes connexes.

En effet, si Φ_n possédait deux zéros (éventuellement confondus) x_i et x_{i+1} dans un tel intervalle, Φ_n serait orthogonal au polynôme $\frac{\Phi_n(x)}{(x - x_i)(x - x_{i+1})}$, donc,

$$0 = \int_S \Phi_n(x) \frac{\Phi_n(x)}{(x - x_i)(x - x_{i+1})} d\mu(x) = \int_S \frac{\Phi_n^2(x)}{(x - x_i)(x - x_{i+1})} d\mu(x),$$

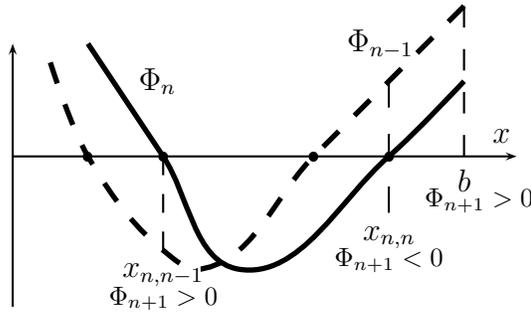
impossible, puisque $(x - x_i)(x - x_{i+1})$ est positif dans S (ce produit n’est négatif que dans l’intervalle $[x_i, x_{i+1}]$ situé hors du support S).

Théorème. Les zéros de Φ_{n+1} sont séparés par ceux de Φ_n , c’est-à-dire que les zéros $x_{1,n} < \dots < x_{n,n}$ de Φ_n et les zéros $x_{1,n+1} < \dots < x_{n+1,n+1}$ de Φ_{n+1} vérifient

$$a < x_{1,n+1} < x_{1,n} < x_{2,n+1} < \dots < x_{n,n+1} < x_{n,n} < x_{n+1,n+1} < b.$$

En effet¹¹, prenons d’abord $n = 1$. Par la récurrence (63), Φ_2 est négatif au zéro $x_{1,1}$ de Φ_1 , puisque $\Phi_0 = 1 > 0$, et Φ_2 est positif en a et b , puisque $\Phi_2(x) = x^2 + \dots$ est positif quand $\pm x$ est grand et que Φ_2 ne peut plus changer de signe quand $x > b$ ou $x < a$. Donc, Φ_2 doit avoir un zéro entre a et $x_{1,1}$ et un autre zéro entre $x_{1,1}$ et b .

¹¹Pour les “experts”: $\{\Phi_{n+1}, \Phi_n, \dots, \Phi_0\}$ est une suite de Sturm. Voir aussi plus loin, § 3.5, p. 94, la même propriété des valeurs propres d’une matrice tridiagonale.



Ensuite, avec le choix de l'unité pour le coefficient de x^n : $\Phi_n(x) = x^n + \dots = \prod_{k=1}^n (x - x_{n,k})$, supposons que les zéros de Φ_n séparent ceux de Φ_{n-1} et montrons que les zéros de Φ_{n+1} séparent bien ceux de Φ_n . Partons de b : $\Phi_{n+1}(b) > 0$ puisque $\Phi_{n+1}(x) = x^{n+1} + \dots$ est le produit des $x - x_{n+1,k}$ tous positifs dès que $x > x_{n+1,n+1}$; $\Phi_{n+1}(x_{n,n}) < 0$ puisque Φ_{n-1} est encore positif ($x_{n-1,n-1} < x_{n,n}$) et que Φ_{n+1} et Φ_{n-1} ont

des signes opposés aux zéros de Φ_n (par la récurrence (63), $\Phi_{n+1}(x) = -\beta_n^2 \Phi_{n-1}(x)$ si $\Phi_n(x) = 0$), donc Φ_{n+1} a au moins un zéro entre $x_{n,n}$ et b ; $\Phi_{n+1}(x_{n-1,n}) > 0$ car Φ_{n-1} a changé de signe entre $x_{n-1,n}$ et $x_{n,n}$, donc Φ_{n+1} a au moins un autre zéro entre $x_{n-1,n}$ et $x_{n,n}$; etc. On trouve ainsi $n + 1$ intervalles

$$(a, x_{1,n}), (x_{1,n}, x_{2,n}), \dots, (x_{n-1,n}, x_{n,n}), (x_{n,n}, b)$$

où Φ_{n+1} de degré $n + 1$ doit s'annuler, il y a donc exactement un zéro dans chacun de ces intervalles. \square

Autre démonstration: Φ_{n+1}/Φ_n est une fonction croissante entre deux asymptotes. En effet, vrai si $n = 0$; ensuite, par la récurrence (63), $\Phi_{n+1}(x)/\Phi_n(x) = x - \alpha_n - \frac{\beta_n^2}{\Phi_n(x)/\Phi_{n-1}(x)}$ a donc un zéro entre deux zéros de Φ_n (pôles).

Exemple. Zéros de quelques polynômes de Legendre sur $[a, b]$.

Rappelons qu'on a trouvé $\Phi_1(x) = x - (a + b)/2$, $\Phi_2(x) = [x - (a + b)/2]\Phi_1(x) - (b - a)^2/12 = [x - (a + b)/2]^2 - (b - a)^2/12$, $\Phi_3(x) = [x - (a + b)/2]\Phi_2(x) - \Phi_1(x) (b - a)^2/15 = [x - (a + b)/2]\{[x - (a + b)/2]^2 - 3(b - a)^2/20\}$:

$$\begin{array}{l} 1 \qquad \qquad \qquad \frac{a+b}{2} \\ 2 \qquad \qquad \qquad \frac{a+b}{2} - \frac{b-a}{\sqrt{12}} \qquad \frac{a+b}{2} + \frac{b-a}{\sqrt{12}} \\ 3 \quad \frac{a+b}{2} - \frac{b-a}{\sqrt{20/3}} \qquad \frac{a+b}{2} \qquad \frac{a+b}{2} + \frac{b-a}{\sqrt{20/3}} \end{array}$$

3.5. Zéros de polynômes orthogonaux et valeurs propres de matrices tridiagonales symétriques.

Prenons maintenant les polynômes orthonormaux, et récrivons la récurrence (65) sous la forme

$$\begin{bmatrix} \alpha_0 & \beta_1 & & & & \\ \beta_1 & \alpha_1 & \beta_2 & & & \\ & & \ddots & \ddots & & \\ & & & \beta_{n-2} & \alpha_{n-2} & \beta_{n-1} \\ & & & & \beta_{n-1} & \alpha_{n-1} \end{bmatrix} \begin{bmatrix} \varphi_0(x) \\ \varphi_1(x) \\ \vdots \\ \varphi_{n-2}(x) \\ \varphi_{n-1}(x) \end{bmatrix} = x \begin{bmatrix} \varphi_0(x) \\ \varphi_1(x) \\ \vdots \\ \varphi_{n-2}(x) \\ \varphi_{n-1}(x) \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \beta_n \varphi_n(x) \end{bmatrix} \tag{67}$$

$$\mathbf{T}_n \boldsymbol{\varphi}_n(x) = x \boldsymbol{\varphi}_n(x) - \beta_n \boldsymbol{\Psi}_n(x).$$

Les zéros $x_{1,n}, \dots, x_{n,n}$ de φ_n sont donc les valeurs propres de la matrice tridiagonale symétrique réelle \mathbf{T}_n . Le vecteur propre correspondant à la valeur propre $x_{j,n}$ est $[\varphi_0(x_{j,n}), \dots, \varphi_{n-1}(x_{j,n})]$. Ceci est très utile pour le calcul (et la théorie) des spectres de matrices symétriques réelles car on peut toujours ramener une telle matrice par un nombre fini de similitudes orthogonales simples à une matrice tridiagonale.

Rappel. Matrices orthogonales et unitaires.

- Une **matrice orthogonale** est une matrice carrée \mathbf{A} qui vérifie

$$\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A} = \mathbf{I}.$$

Les lignes d'une telle matrice sont donc de norme unité et sont orthogonales entre elles. Les colonnes d'une telle matrice sont donc de norme unité et sont orthogonales entre elles (au sens de l'orthogonalité des vecteurs de \mathbb{R}^N : $(\varphi, \psi) = \varphi^T \psi = \sum_{k=1}^N \varphi_k \psi_k$).

- Une **matrice unitaire** est une matrice carrée \mathbf{A} qui vérifie

$$\mathbf{A}\mathbf{A}^H = \mathbf{A}^H\mathbf{A} = \mathbf{I}.$$

Les lignes d'une telle matrice sont donc de norme unité et sont orthogonales entre elles. Les colonnes d'une telle matrice sont donc de norme unité et sont orthogonales entre elles (au sens de l'orthogonalité des vecteurs de \mathbb{C}^N : $(\varphi, \psi) = \varphi^H \psi = \sum_{k=1}^N \overline{\varphi_k} \psi_k$).

- Toute matrice symétrique réelle est diagonalisable par similitude orthogonale réelle:

$$\mathbf{A} = \mathbf{A}^T \Rightarrow \mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}, \quad \mathbf{V}^{-1} = \mathbf{V}^T.$$

Les colonnes de \mathbf{V} forment une base orthonormale (au sens de \mathbb{R}^N) de vecteurs propres de \mathbf{A} .

- Toute matrice hermitienne est diagonalisable par similitude unitaire:

$$\mathbf{A} = \mathbf{A}^H \Rightarrow \mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}, \quad \mathbf{V}^{-1} = \mathbf{V}^H.$$

Les colonnes de \mathbf{V} forment une base orthonormale (au sens de \mathbb{C}^N) de vecteurs propres de \mathbf{A} .

Soit $x_{j,n}$ le $j^{\text{ème}}$ zéro de φ_n . Par (67), la $j^{\text{ème}}$ colonne de \mathbf{V} a pour composantes $K_j \varphi_0(x_{j,n}), K_j \varphi_1(x_{j,n}), \dots, K_j \varphi_{n-1}(x_{j,n})$, où K_j est tel que la somme des carrés de ces composantes soit égale à 1: $K_j = [\sum_{k=0}^{n-1} \varphi_k^2(x_{j,n})]^{-1/2}$. L'élément i, j de \mathbf{V} est donc

$$v_{i,j} = \frac{\varphi_i(x_{j,n})}{\sqrt{\sum_{k=0}^{n-1} \varphi_k^2(x_{j,n})}}, \quad i = 0, \dots, n-1, j = 1, \dots, n.$$

Exprimons maintenant que les *lignes* de \mathbf{V} sont orthonormales (au sens de \mathbb{R}^n):

Théorème. *Les polynômes orthonormaux $\varphi_0, \varphi_1, \dots, \varphi_{n-1}$ selon un produit scalaire $(f, g) = \int_a^b f(x)g(x)d\mu(x)$, où μ a au moins n points de croissance sur $[a, b]$, sont également orthonormaux selon le produit scalaire discret*

$$(f, g)_n = \sum_{j=1}^n w_{j,n} f(x_{j,n}) g(x_{j,n}),$$

où $x_{1,n}, \dots, x_{n,n}$ sont les zéros de Φ_n , et les poids $w_{j,n}$ sont

$$w_{j,n} = \frac{1}{\sum_{k=0}^{n-1} \varphi_k^2(x_{j,n})}, \quad j = 1, \dots, n.$$

En effet, l'orthonormalité (dans \mathbb{R}^n) des lignes de \mathbf{V} donne bien

$$\sum_{j=0}^n v_{i_1,j} v_{i_2,j} = \sum_{j=1}^n \frac{\varphi_{i_1}(x_{j,n}) \varphi_{i_2}(x_{j,n})}{\sum_{k=0}^{n-1} \varphi_k^2(x_{j,n})} = \delta_{i_1,i_2},$$

pour $0 \leq i_1, i_2 \leq n-1$. □

Voici une très importante application de cette théorie:

3.6. Formules d'intégration de Gauss. Première approche.

D'après le théorème précédent, le produit scalaire initial ne se distingue pas du produit scalaire discret à n points tant qu'on se contente d'effectuer des produits scalaires d'éléments de \mathcal{P}_{n-1} entre eux. Comme le produit scalaire utilise le produit des deux fonctions considérées, le produit scalaire discret permet le calcul exact d'intégrales $\int_a^b f(x) d\mu(x) = (f, 1)$ pour $\forall f \in \mathcal{P}_{2n-2}$, en fait \mathcal{P}_{2n-1} : soit f un polynôme de degré $\leq 2n-1$, divisons le par φ_n , et développons le quotient et le reste (tous deux dans \mathcal{P}_{n-1}) dans la base $\{\varphi_0, \dots, \varphi_{n-1}\}$:

$$f = [\xi_0 \varphi_0 + \dots + \xi_{n-1} \varphi_{n-1}] \varphi_n + \eta_0 \varphi_0 + \dots + \eta_{n-1} \varphi_{n-1},$$

$$(f, 1) = \xi_0(\varphi_0, \varphi_n) + \dots + \xi_{n-1}(\varphi_{n-1}, \varphi_n) + \frac{\eta_0(\varphi_0, \varphi_0) + \dots + \eta_{n-1}(\varphi_{n-1}, \varphi_0)}{\varphi_0} = \frac{\eta_0}{\varphi_0},$$

$$(f, 1)_n = \xi_0(\varphi_0, \varphi_n)_n + \dots + \xi_{n-1}(\varphi_{n-1}, \varphi_n)_n + \frac{\eta_0(\varphi_0, \varphi_0)_n + \dots + \eta_{n-1}(\varphi_{n-1}, \varphi_0)_n}{\varphi_0} = \frac{\eta_0}{\varphi_0},$$

les deux intégrales coïncident donc, les produits scalaires étant égaux dans les deux cas (les produits scalaires originaux de φ_n et d'un polynôme de degré $\leq n-1$ sont nuls par orthogonalité, les produits scalaires discrets de φ_n et de n'importe quelle fonction sont nuls parce que tous les $\varphi_n(x_{j,n}) = 0$). On a donc

$$\forall f \in \mathcal{P}_{2n-1}, \quad \int_a^b f(x) d\mu(x) = \sum_{j=1}^n w_{j,n} f(x_{j,n}).$$

Inversement, si on se propose d'approcher $\int_a^b f(x) d\mu(x)$ par une somme pondérée $\sum_{j=1}^n H_j f(x_j)$, la formule devant être exacte pour des polynômes de degré aussi élevé que possible (*degré de précision de la formule*), on retrouve la même formule. Comme il y a $2n$ paramètres H_1, \dots, H_n et x_1, \dots, x_n à déterminer, on peut espérer réaliser un degré de précision de $2n-1$, puisqu'il y a $2n$ coefficients dans un polynôme de degré $\leq 2n$:

$$\forall f \in \mathcal{P}_{2n-1}, \quad \int_a^b f(x) d\mu(x) = \sum_{j=1}^n H_j f(x_j).$$

Exprimons cela pour $f(x) = 1, x, \dots, x^{2n-1}$:

$$\begin{aligned} \mu_0 &= H_1 & + & H_2 & + \dots & + & H_n \\ \mu_1 &= H_1 x_1 & + & H_2 x_2 & + \dots & + & H_n x_n \\ \mu_2 &= H_1 x_1^2 & + & H_2 x_2^2 & + \dots & + & H_n x_n^2 \\ \dots & & & & & & \dots \\ \mu_{2n-1} &= H_1 x_1^{2n-1} & + & H_2 x_2^{2n-1} & + \dots & + & H_n x_n^{2n-1} \end{aligned}$$

Redoutable (apparemment) système de $2n$ équations non linéaires pour les $2n$ inconnues $H_1, \dots, H_n, x_1, \dots, x_n$.


```

6      0.6931471798865279786405606852820113472021359...
7      0.6931471805400134518743042766974180037770983...
8      0.6931471805593561216771329329612392432031989...
9      0.6931471805599279082472626955838961065402593...
10     0.6931471805599447958100734473749544043869926...
...
15     0.6931471805599453094172206753728118939529066...
20     0.6931471805599453094172321214579224966600492...
25     0.6931471805599453094172321214581765680698696...
30     0.6931471805599453094172321214581765680755001...
31     0.6931471805599453094172321214581765680755001...

```

étonnant, non?

From: Bill McLean <w.mclean@unsw.edu.au>
Date: Mon, 14 May 2007 03:56:38 -0400
Subject: subject: f95 version of gaussq.f

The fortran 95 module gaussquad.f95 can compute the integration points and weights for any of the classical Gauss quadrature rules, including the Radau and Lobatto variants. Included is a low-level routine that can handle any Gauss rule given the coefficients in the 3-term recurrence relation for the associated orthogonal polynomials. You can download the source distribution from

<http://www.maths.unsw.edu.au/~mclean/gaussquad-2.4.tgz>

My f95 code is a goto-less version of the f77 routines in netlib's gaussq.f, and should produce nearly identical results. The new package adds an optional C language interface that relies on a couple of the interoperability features of the f2003 standard, plus a few test programs and examples in fortran, C and python.

3.7. Formule d'intégration de Gauss et fractions continues.

Soit S une partie de \mathbb{R} , $z \notin S$, $F_n(z) := \sum_1^n \frac{H_i}{z - x_{i,n}}$ est donc une approximation de $F(z) := \int_S \frac{d\mu(x)}{z - x}$, *transformée de Stieltjes* de $d\mu$. Pas mal de fonctions très importantes sont des transformées de Stieltjes, ainsi, si $d\mu(x) = dx$ sur $S = [a, b]$, $F(z) = \log((z - a)/(z - b))$.

$F_n(z)$ est une fonction rationnelle de z de dénominateur $\prod_1^n (z - x_{i,n}) = \Phi_n(z)$, on a donc $F_n(z) = \frac{\Psi_n(z)}{\Phi_n(z)}$, où $\Psi_n \in \mathcal{P}_{n-1}$.

On a (cf. (63), p. 88) la récurrence $\Phi_{n+1}(x) = (x - \alpha_n) \Phi_n(x) - \beta_n^2 \Phi_{n-1}(x)$. Qu'en est-il de Ψ_n ? Amorçons le développement en série de puissances de $1/z$ de $F(z)$ par

$$\begin{aligned}
F(z) &= \int_S \frac{1 - x^{p+1}/z^{p+1}}{z(1 - x/z)} d\mu(x) + \int_S \frac{x^{p+1}/z^{p+1}}{z - x} d\mu(x) \\
&= \int_S \left(\frac{1}{z} + \frac{x}{z^2} + \frac{x^2}{z^3} + \cdots + \frac{x^p}{z^{p+1}} \right) d\mu(x) + \int_S \frac{x^{p+1}}{z^{p+1}(z - x)} d\mu(x) \\
&= \frac{\mu_0}{z} + \frac{\mu_1}{z^2} + \frac{\mu_2}{z^3} + \cdots + \frac{\mu_p}{z^{p+1}} + O(z^{-p-2})
\end{aligned}$$

Prenons $p = 2n - 1$. Comme la formule de Gauss à n points appliquée à $1, x, \dots, x^{2n-1}$ coïncide avec le moment $\mu_0, \mu_1, \dots, \mu_{2n-1}$, on a aussi $F_n(z) = \frac{\mu_0}{z} + \frac{\mu_1}{z^2} + \frac{\mu_2}{z^3} + \dots + \frac{\mu_{2n-1}}{z^{2n}} + O(z^{-2n-1})$, on a

$$F(z) - F_n(z) = O(z^{-2n-1}),$$

ou $\Phi_n(z)(F(z) - F_n(z)) = \Phi_n(z)F(z) - \Psi_n(z) = O(z^{-n-1})$. Remarquons que $\Psi_0 = 0$.

Examinons maintenant

$$\begin{aligned} & \Phi_{n+1}(z)F(z) - \Psi_{n+1}(z) - (z - \alpha_n)[\Phi_n(z)F(z) - \Psi_n(z)] + \beta_n^2[\Phi_{n-1}(z)F(z) - \Psi_{n-1}(z)] \\ &= -[\Psi_{n+1}(z) - (z - \alpha_n)\Psi_n(z) + \beta_n^2\Psi_{n-1}(z)] \end{aligned}$$

quand $n \geq 1$. Cette expression tend vers zéro (au moins aussi vite que z^{-n}) quand $z \rightarrow \infty$; mais comme elle est aussi un polynôme, elle doit être nulle: les Ψ_n vérifient la même récurrence que les Φ_n !

Exercice. Montrez que, si $\Phi_n = T_n$, et $S = [-1, 1]$, les H_i sont tous égaux à π/n (formule de **Gauss-Tchebycheff**), que $F(z) = \pi(z^2 - 1)^{-1/2}$, $\Psi_n = U_{n-1}$.

On a donc $\lambda\Phi_{m+1}(z) - \Psi_{m+1}(z) - (z - \alpha_m)[\lambda\Phi_m(z) - \Psi_m(z)] + \beta_m^2[\lambda\Phi_{m-1}(z) - \Psi_{m-1}(z)] = 0$, $m \geq 1$, quel que soit λ .

En $m = 0$, on a $\lambda\Phi_1(z) - \Psi_1(z) - (z - \alpha_0)[\lambda\Phi_0(z) - \Psi_0(z)] = -\Psi_1 = -\mu_0$.

Choisissons $\lambda = F_n(z) = \Psi_n(z)/\Phi_n(z)$ pour z fixé hors de $[a, b]$. Soit $\chi_m = F_n(z)\Phi_m(z) - \Psi_m(z)$ et $\rho_m = \chi_m/\chi_{m-1}$. On a donc $\chi_{m+1} - (z - \alpha_m)\chi_m + \beta_m^2\chi_{m-1} = 0$, $m = 1, \dots, n - 1$.

En $m = n - 1$, on a $\chi_n = 0$, donc $\rho_{n-1} = \frac{\beta_{n-1}^2}{z - \alpha_{n-1}}$, puis

$$\begin{aligned} \rho_m &= \frac{\chi_m}{\chi_{m-1}} = \frac{\beta_m^2\chi_m}{\beta_m^2\chi_{m-1}} = \frac{\beta_m^2\chi_m}{(z - \alpha_m)\chi_m - \chi_{m+1}} = \frac{\beta_m^2}{z - \alpha_m - \frac{\beta_{m+1}^2}{z - \alpha_{m+1} - \frac{\beta_{m+2}^2}{z - \alpha_{m+2} - \dots - \frac{\beta_{n-1}^2}{z - \alpha_{n-1}}}}} \\ &= \frac{\beta_m^2}{z - \alpha_m - \rho_{m+1}} \end{aligned}$$

Enfin, en $m = 0$, $\chi_1 - (z - \alpha_0)\chi_0 = -\mu_0$, donc $\rho_1 - (z - \alpha_0) = -\mu_0/\chi_0 = -\mu_0/F_n(z)$, d'où

$$F_n(z) = \frac{\mu_0}{z - \alpha_0 - \rho_1} = \frac{\mu_0}{z - \alpha_0 - \frac{\beta_1^2}{z - \alpha_1 - \frac{\beta_2^2}{z - \alpha_2 - \frac{\beta_3^2}{z - \alpha_3 - \dots - \frac{\beta_{n-1}^2}{z - \alpha_{n-1}}}}}$$

que l'on peut aussi voir à partir de la résolution par triangularisation de

$$\begin{bmatrix} z - \alpha_0 & -1 & & & & \\ -\beta_1^2 & z - \alpha_1 & -1 & & & \\ & & \ddots & \ddots & & \\ & & & & \ddots & \\ & & & & & -\beta_{n-1}^2 & z - \alpha_{n-1} \end{bmatrix} \begin{bmatrix} \chi_0 \\ \chi_1 \\ \vdots \\ \chi_{n-1} \end{bmatrix} = \begin{bmatrix} \mu_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} .$$

en procédant aux éliminations à partir de la dernière équation. On a aboutit à une matrice bidiagonale triangulaire inférieure dont les éléments diagonaux sont β_m^2/ρ_m , $m = 0, \dots, n - 1$.

Peut-on donner un sens à $\frac{\mu_0}{z - \alpha_0 - \frac{\beta_1^2}{z - \alpha_1 - \frac{\beta_2^2}{z - \alpha_2 - \dots}}}$ comme limite de $F_n(z)$ quand $n \rightarrow \infty$?

Théorème de Markoff. Si $S \subset [a, b]$, avec $-\infty < a < b < \infty$, $F_n(z)$ tend vers $F(z)$ quand $n \rightarrow \infty$, uniformément en z dans tout compact de $\mathbb{C} \setminus [a, b]$.

En effet, comme la formule d'intégration de Gauss à n points est exacte pour tout polynôme de degré $\leq 2n - 1$, soit $p(x; z)$ une approximation dans \mathcal{P}_{2n-1} de $1/(z - x)$.

On a $\int_S p(x; z) d\mu(x) = \sum_1^n H_i p(x_{i,n}; z)$,

$$F(z) - F_n(z) = \int_S \left(\frac{1}{z-x} - p(x; z) \right) d\mu(x) - \sum_1^n H_i \left(\frac{1}{z-x_{i,n}} - p(x_{i,n}; z) \right).$$

Soit $\varepsilon_{2n-1}(z) = \|(z-x)^{-1} - p(x; z)\|_\infty$ sur $x \in [a, b]$ (nous savons que les $x_{i,n} \in [a, b]$). Alors, $|F(z) - F_n(z)| \leq \varepsilon_{2n-1}(z) [\int_S d\mu(x) + \sum_1^n H_i] = 2\mu_0 \varepsilon_{2n-1}(z)$.

On estime $\varepsilon_{2n-1}(z)$ avec précision par $2(b-a)^{-1} |(\zeta - \xi)^{-1} - \sum_0^{2n-1} c_k T_k(\xi)| \leq 2(b-a)^{-1} \sum_{2n}^\infty |c_k| = 8(b-a)^{-1} |(1-p)^{-1}(1-p^2)^{-1} p^{2n+1}|$, où $x = (a+b)/2 + (b-a)\xi/2$, $z = (a+b)/2 + (b-a)\zeta/2$, et $p = \zeta \pm (\zeta^2 - 1)^{1/2} = (2z - a - b \pm 2[(z-a)(z-b)]^{1/2}) / (b-a)$ tel que $|p| < 1$ (cf. table p. 59: $c_k = 4p^{k+1}/(1-p^2)$, où ζ joue ici le rôle de c).

Exercice. Montrez que $F(z) = \frac{\beta}{z - \alpha - \frac{\beta^2}{z - \alpha - \frac{\beta^2}{\ddots}}}$ = $p(z) = (2z - a - b \pm 2[(z-a)(z-b)]^{1/2}) / (b-a)$, avec

$a = \alpha - 2\beta$, $b = \alpha + 2\beta$, $|p(z)| < 1$ quand $z \in \mathbb{C} \setminus [a, b]$. On a $\Phi_n(z) = \beta^n U_n \left(\frac{z-\alpha}{2\beta} \right)$, $\Psi_n(z) = \beta \Phi_{n-1}(z)$, $d\mu(x) = [(b-x)(x-a)]^{1/2} / (2\beta\pi)$ sur $S = [a, b]$. Bien sûr, on s'aperçoit bien formellement que $F(z) = \frac{\beta}{z - \alpha - \beta F(z)}$, donc que $F(z)$ est une racine de $\beta\chi^2 - (z - \alpha)\chi + \beta = 0$, mais il n'est pas si simple de savoir laquelle choisir, ni quand la fraction continue converge.

3.8. Formule de Christoffel-Darboux.

Multiplions (67) à gauche par le vecteur ligne $\varphi_n(y)^T = [\varphi_0(y), \dots, \varphi_{n-1}(y)]$:

$$\varphi_n(y)^T \mathbf{T}_n \varphi_n(x) = x \varphi_n(y)^T \varphi_n(x) - \beta_n \varphi_{n-1}(y) \varphi_n(x),$$

permutons x et y , et soustrayons:

$$\varphi_n(y)^T \varphi_n(x) = \sum_{i=0}^{n-1} \varphi_i(x) \varphi_i(y) = \beta_n \frac{\varphi_n(x) \varphi_{n-1}(y) - \varphi_n(y) \varphi_{n-1}(x)}{x - y}$$

On verra l'importance de cette formule à propos des *polynômes noyaux*.

Si on fait tendre y vers x :

$$\sum_{i=0}^{n-1} \varphi_i(x)^2 = \beta_n [\varphi_n'(x) \varphi_{n-1}(x) - \varphi_n(x) \varphi_{n-1}'(x)]$$

Exercice. Montrez que $S = \sum_{k=0}^n U_k(x) U_{n-k}(y) = \frac{U_{n+1}(y) - U_{n+1}(x)}{2(y-x)}$.

On multiplie $2x$ $\begin{bmatrix} U_0(x) \\ \vdots \\ U_n(x) \end{bmatrix} = \mathbf{T}_{n+1} \begin{bmatrix} U_0(x) \\ \vdots \\ U_n(x) \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ U_{n+1}(x) \end{bmatrix}$ à gauche par le vecteur ligne $[U_n(y), \dots, U_0(y)]$

pour obtenir $2xS = T(x, y) + U_{n+1}(x)$ et on s'aperçoit que $T(x, y) = [U_n(y), \dots, U_0(y)] \begin{bmatrix} 0 & 1 & & \\ 1 & 0 & 1 & \\ & \ddots & \ddots & \ddots \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} U_0(x) \\ \vdots \\ U_n(x) \end{bmatrix}$

est symétrique en x et en y .

3.9. Orthogonalité et opérateurs (formellement) hermitiens.

Les fonctions propres d'opérateurs formellement hermitiens fournissent d'intéressantes classes de fonctions orthogonales. D'abord, quelques définitions:

Définition. *Un opérateur \mathcal{A} dans un espace préhilbertien X (c'est-à-dire une application linéaire définie sur un sous-espace D de X et à valeurs dans X) est **formellement hermitien**¹⁴ si*

$$\forall f, g \in D, \quad (\mathcal{A}f, g) = (f, \mathcal{A}g). \quad (68)$$

L'exemple le plus simple est fourni par les matrices carrées symétriques sur \mathbb{R}^N , ou hermitiennes sur \mathbb{C}^N , munis du produit scalaire usuel.

Et voici où apparaît l'orthogonalité:

Propositions. *Les valeurs propres d'un opérateur formellement hermitien sont réelles. Deux vecteurs propres d'un opérateur formellement hermitien correspondant à deux valeurs propres différentes sont orthogonaux.*

Ce ne sont là que deux tout petits fragments de la théorie spectrale des opérateurs: si $\mathcal{A}f = \lambda f$, $(\mathcal{A}f, f) = \lambda \|f\|^2 = (f, \mathcal{A}f) = \overline{\lambda} \|f\|^2$; si f et g sont deux vecteurs propres de \mathcal{A} correspondant aux deux valeurs propres λ et μ , $(\mathcal{A}f, g) = \lambda(f, g) = (f, \mathcal{A}g) = \mu(f, g)$, d'où $(f, g) = 0$ si $\lambda \neq \mu$. \square

Proposition. *Soit w une densité sur (a, b) , et le produit scalaire $(f, g)_w = \int_a^b f(x)g(x) w(x) dx$.*

L'opérateur différentiel du second ordre

$$L = p(x) \frac{d^2}{dx^2} + q(x) \frac{d}{dx} + r(x) \quad (69)$$

est formellement hermitien sur tout sous-espace de $\mathcal{C}^2(a, b)$ de fonctions vérifiant $(f, f)_w < \infty$ et $\lim_{\substack{x \rightarrow a, b \\ x \in (a, b)}} p(x)w(x)f(x) = 0$ si

$$\frac{d(p(x)w(x))}{dx} = q(x)w(x), \quad \text{ou} \quad \frac{w'(x)}{w(x)} = \frac{q(x) - p'(x)}{p(x)} \quad a < x < b. \quad (70)$$

En effet,

$$\begin{aligned} (Lf, g)_w &= \int_a^b [pf'' + qf' + rf]gw dx = [pgwf']_a^b + \int_a^b \{[-(pw)' + qw]gf' - pwg'f' + rfg\} dx \\ (f, Lg)_w &= \int_a^b f[pqg'' + qg' + rg]w dx = [fpwg']_a^b + \int_a^b \{[-(pw)' + qw]fg' - pwf'g' + rfg\} dx \end{aligned}$$

sont égaux si (70) est vrai, et si $f(x)p(x)w(x)$ et $g(x)p(x)w(x)$ tendent vers 0 quand $x \rightarrow a$ ou b . \square

¹⁴Pour avoir le droit d'être appelé *hermitien*, un opérateur doit avoir un domaine de définition D dense dans un espace de Hilbert séparable [on verra cela plus loin] et vérifier (68). La théorie spectrale des opérateurs hermitiens prépare celle des **opérateurs autoadjoints** (Cf. J. Dieudonné, *Éléments d'analyse*, II, Gauthier-Villars, 1969, § 15.13).

La relation entre *formellement hermitien* et *autoadjoint* est un peu la relation qu'il y a entre *préhilbertien* et *hilbertien*. . . On y reviendra plus tard.

Remarques: si $w = 1$, Lf s'écrit commodément

$$(pf')' + rf \tag{71}$$

(opérateur de **Sturm-Liouville**)¹⁵. En général, on a ici $wLf = (wpf')' + wrf$.

Un opérateur de Sturm-Liouville est dit **singulier** si $w(a)p(a) = w(b)p(b) = 0$. Les conditions d'annulation aux limites sont alors automatiquement vérifiées si le domaine de définition de l'opérateur est restreint aux fonctions continûment dérivables sur $[a, b]$.

Tout opérateur de Sturm-Liouville fournit donc automatiquement des fonctions propres orthogonales très utiles (nombreux exemples en physique et en mécanique). Les cas où on retrouve des *polynômes* sont parfaitement circonscrits:

3.10. Polynômes orthogonaux classiques.

Théorème de Bochner¹⁶. Une famille de polynômes orthogonaux $\{\Phi_n\}_0^\infty$ est l'ensemble des fonctions propres d'un opérateur de la forme (69) seulement dans les trois cas suivants (**polynômes orthogonaux classiques**):

$p(x)$	$w(x)$	nom
$(x - a)(b - x)$	$C(b - x)^\alpha(x - a)^\beta, \quad a < x < b; \quad \alpha, \beta > -1$	Jacobi
$x - a$	$C(x - a)^\alpha e^{-Ax}, \quad a < x < \infty; \quad \alpha > -1, \text{ si } A > 0$ $, \quad -\infty < x < a; \quad \alpha > -1, \text{ si } A < 0$	Laguerre
1	$Ce^{-Ax^2 - Bx}, \quad -\infty < x < \infty; \quad A > 0$	Hermite

Les polynômes Φ_n sont alors des solutions particulières des équations différentielles linéaires du second ordre

$$p(x)\Phi_n''(x) + q(x)\Phi_n'(x) + r\Phi_n(x) = \lambda_n\Phi_n(x) \tag{72}$$

où $p \in \mathcal{P}_2, q \in \mathcal{P}_1$ ¹⁷, $r = \text{constante}$.

Démonstration. Nous savons déjà que les fonctions propres de (69) sont orthogonales selon $(\cdot, \cdot)_w$ si w vérifie (70).

Que peut-on dire de p, q et r ? Ne considérons que Φ_0, Φ_1 et Φ_2 :

$$\Phi_0 \in \mathcal{P}_0 \Rightarrow r\Phi_0 = \lambda_0\Phi_0 \Rightarrow r \in \mathcal{P}_0,$$

$$\Phi_1 \in \mathcal{P}_1 \Rightarrow q\Phi_1' + r\Phi_1 = \lambda_1\Phi_1 \Rightarrow q \in \mathcal{P}_1,$$

$$\Phi_2 \in \mathcal{P}_2 \Rightarrow p\Phi_2'' + q\Phi_2' + r\Phi_2 = \lambda_2\Phi_2 \Rightarrow p \in \mathcal{P}_2.$$

(N.B. Φ_n doit être de degré exact n).

Discutons (70):

¹⁵Attention, on appelle parfois *formellement autoadjoints* des opérateurs différentiels simplement donnés par une formule de type (71) sans autre précision quant au domaine de définition.

¹⁶S. Bochner, Über Sturm-Liouvillesche Polynomsysteme, *Math. Zeit.* **29** (1929) 730-736, cité, ainsi que E.J. Routh, On some properties of certain solutions of a differential equation of the second order, *Proc. London Math. Soc.* **16** (1885) 245-261, par W.A. Al-Salam, Characterization theorems for orthogonal polynomials, pp. 1-24 in *Orthogonal Polynomials: Theory and Practice*, (P. Nevai, ed.), NATO ASI Series C: Math. and Phys. Sci. **294**, Kluwer, 1990.

¹⁷**Exercice:** en fait, q est de degré exact 1! (A. Ronveaux)

- (1) p a deux zéros distincts \tilde{a} et \tilde{b} . Le développement du membre de droite de (70) en fractions simples donne $\alpha/(x-\tilde{b}) + \beta/(x-\tilde{a})$, d'où $w(x) = \text{constante } (x-\tilde{b})^\alpha (x-\tilde{a})^\beta$. La densité w doit être de signe constant dans (a, b) , et il faut $p(x)w(x) \rightarrow 0$ quand $x \rightarrow a$ ou b , d'où $\tilde{a} = a$ et $\tilde{b} = b$, et aussi α et $\beta > -1$. Ces deux dernières inégalités assurent d'ailleurs l'existence de moments finis.
- (2) p a un zéro double: on ne trouve pas d'intervalle d'orthogonalité dans ce cas¹⁸.
- (3) p est du premier degré. On a alors $(q-p')/p = -A + \alpha/(x-\tilde{a})$, donc $w(x) = \text{constante } (x-\tilde{a})^\alpha \exp(-Ax)$. Signe constant de w et $\lim pw = 0$ en a et $b \Rightarrow a = \tilde{a}$ et $b = +\infty$ si $A > 0$, l'intervalle d'orthogonalité est $(-\infty, a)$ si $A < 0$.
- (4) p est constant. Alors, soit $(q-p')/p = -2Ax - B$: $w(x) = \text{constante } \exp(-Ax^2 - Bx)$ sur $(-\infty, \infty)$. A doit être strictement positif.

En examinant la contribution de x^n à (72), on obtient

$$\lambda_n = p'' \frac{n(n-1)}{2} + q'n + r. \quad (73)$$

Suite de la démonstration. Jusqu'ici, on n'a encore trouvé que trois fonctions propres Φ_0, Φ_1 et Φ_2 .

Proposition: Si $p \in \mathcal{P}_2, q \in \mathcal{P}_1, r \in \mathcal{P}_0$, l'équation $py'' + qy' + ry = \lambda_n y$ admet exactement une solution polynomiale de degré n si $\lambda_n = p''n(n-1)/2 + q'n + r$, et si $-2q'/p''$ n'est pas un entier positif impair.

1. Par identification de coefficients.

Prenons n entier positif quelconque et appliquons (69) à une fonction propre candidate de degré n

$$\Phi_n(x) = \sum_{k=0}^n c_k x^k,$$

avec $c_n = 1$.

Alors, comme p, q et r sont de degrés $\leq 2, 1$ et 0 ,

$$\begin{aligned} p(x)\Phi_n''(x) &= (p_0 + p_1x + p_2x^2) \left(\sum_{k=2}^n c_k k(k-1)x^{k-2} \right), \\ q(x)\Phi_n'(x) &= (q_0 + q_1x) \left(\sum_{k=1}^n c_k kx^{k-1} \right), \\ r\Phi_n(x) &= r \left(\sum_{k=0}^n c_k x^k \right). \end{aligned}$$

Le coefficient de x^k de la somme est $p_0(k+2)(k+1)c_{k+2} + (k+1)(p_1k + q_0)c_{k+1} + (p_2k(k-1) + q_1k + r)c_k$, $k = 0, 1, \dots, n$, où on a posé $c_{n+1} = c_{n+2} = 0$, et qu'il faut évaluer au coefficient de x^k dans $\lambda_n \Phi_n(x)$:

$$p_0(k+2)(k+1)c_{k+2} + (k+1)(p_1k + q_0)c_{k+1} + (p_2k(k-1) + q_1k + r - \lambda_n)c_k = 0, k = 0, 1, \dots, n.$$

¹⁸Mais les Φ_n ont été étudiés (polynômes de Bessel), ils interviennent dans des approximations rationnelles de l'exponentielle.

En $k = n$, il ne reste que le terme en $c_n \neq 0$, il faut donc

$$\lambda_n = p_2 n(n-1) + q_1 n + r.$$

Puis, en $k = n-1, n-2, \dots, 0$ on a une récurrence qui donne $c_{n-1}, c_{n-2}, \dots, c_0$ pourvu que le coefficient de c_k soit non nul, c'est-à-dire

$$p_2 k(k-1) + q_1 k + r - \lambda_n = (k-n)(p_2(k+n-1) + q_1) \neq 0.$$

Le problème ne se pose que si p est effectivement de degré 2 (cas Jacobi), $w(x) = C(b-x)^\alpha(x-a)^\beta$, $w'(x)/w(x) = (\alpha+\beta)/x + \dots = (q(x)-p'(x))/p(x)$ (par (70)) $= (q_1/p_2 - 2)/x + \dots$, donc $q_1/p_2 = \alpha + \beta + 2$ et le problème ne surgit que si $\alpha + \beta = -k - n - 1$, or α et β doivent être strictement supérieurs à -1 (intégrabilité de w), donc le problème ne se pose pas dès que $n > 0$.

Si on développe Φ_n en puissances de $x - a$ ou de $x - b$, on a des relations de récurrence plus simples (développements *hypergéométriques*, voir références p. 108).

2. Par formule de Rodrigues.

Multiplions (72) (nous ne savons pas encore si des solutions polynomiales existent pour tous les degrés) par $w(x)$ et utilisons (70): $(pw)' = qw \Rightarrow$

$$(pw\Phi_n')' = (\lambda_n - r)w\Phi_n.$$

Une structure similaire apparaît si on dérive (72) k fois:

$$p\Phi_n^{(k+2)} + (kp' + q)\Phi_n^{(k+1)} + (k(k-1)p''/2 + kq' + r)\Phi_n^{(k)} = \lambda_n\Phi_n^{(k)},$$

p'' et q' étant des constantes. Donc, toujours par (70) $(pw)' = qw \Rightarrow (p^{k+1}w)' = (q + kp')p^k w$, et en multipliant par $p^k w$

$$(p^{k+1}w\Phi_n^{(k+1)})' = [\lambda_n - k(k-1)p''/2 - kq' - r]p^k w\Phi_n^{(k)}.$$

Donc, $p^k w\Phi_n^{(k)}$ est une constante fois la dérivée première d'une fonction de même type $p^{k+1}w\Phi_n^{(k+1)}$, donc dérivée seconde de $p^{k+2}w\Phi_n^{(k+2)}$, etc. On arrive, avec $k = 0$, à

$$\Phi_n(x) = \text{const.} \frac{1}{w(x)} \frac{d^n w(x)p^n(x)}{dx^n}, \quad (74)$$

une formule trouvée par Benjamin Olinde Rodrigues (1794-1851) pour les polynômes de Legendre ($w(x) \equiv 1$)¹⁹, et par Jacobi (voir p. 51) pour les polynômes de Tchebycheff.

Il reste à vérifier que (74) donne bien un polynôme de degré exact n . En effet, grâce à (70), les dérivées successives de wp^n sont wp^{n-1} fois le polynôme du premier degré $q + (n-1)p'$, wp^{n-2} fois un polynôme de degré 2, etc. \square

3. Par factorisation d'opérateurs (Darboux).

Si on dispose de Φ_n de degré n tel que Φ_n soit fonction propre de (72), montrons comment construire une nouvelle fonction propre Φ_{n+1} par

$$\Phi_{n+1} = p\Phi_n' + s_n\Phi_n, \quad (75)$$

¹⁹O. Rodrigues, Mémoire sur l'attraction des sphéroïdes, *Correspondance sur l'École Polytechnique* **3** (1816) 361-385. Voir aussi Area, I.; Godoy, E.; Ronveaux, A.; Zarzo, A.: Hypergeometric-type differential equations: second kind solutions and related integrals, *J. Comput. Appl. Math.* **157** (2003), no. 1, 93-106; `rodrigues.txt`; André Ronveaux, Jean Mawhin: Rediscovering the contributions of Rodrigues on the representation of special functions, *Expositiones Mathematicæ* **23**, Issue 4, 1 December 2005, Pages 361-369.

où s_n est un polynôme de degré ≤ 1 , en adaptant une technique de Darboux²⁰, que l'on présente actuellement comme suit: si on arrive à factoriser un opérateur différentiel R comme produit de deux opérateurs différentiels $R = MN$, alors, si y est une fonction propre de R : $Ry = \lambda y$, Ny est une fonction propre de NM , de même valeur propre λ . En effet (et cela paraît presque trop simple), $MNy = \lambda y \Rightarrow NMNy = NM(Ny) = \lambda Ny!$

La première idée serait de factoriser L , puisque Φ_n est une fonction propre de L (de valeur propre λ_n), mais cela n'aboutit pas.

Ce qui marche, c'est de (presque) factoriser $p(L - \lambda_n)$:

$$M_n N_n y = \left(p \frac{d}{dx} + \alpha_n \right) \left(p \frac{d}{dx} + \beta_n \right) y = p^2 y'' + p(p' + \alpha_n + \beta_n) y' + (p\beta'_n + \alpha_n \beta_n) y = p(L - \lambda_n) y - \gamma_n y, \quad (76)$$

si $\alpha_n + \beta_n = q - p'$ et $p\beta'_n + \alpha_n \beta_n = p(r - \lambda_n) - \gamma_n$, où γ_n est une constante encore inconnue, tout comme les coefficients de α_n et $\beta_n \in \mathcal{P}_1$ sont encore inconnus. On a une équation de Riccati pour β_n à résoudre dans \mathcal{P}_1 : $p\beta'_n + (q - p')\beta_n - \beta_n^2 = p(r - \lambda_n) - \gamma_n$. Coefficients de x^2 : équation du second degré pour β'_n : $p''\beta'_n/2 + (q' - p'')\beta'_n - (\beta'_n)^2 = p''(-q'n - n(n-1)p''/2)$ (on a utilisé (73)), de racines $\beta_n^{(+)} = p''(n-1)/2 + q'$ et $\beta_n^{(-)} = -p''n/2$. Avec $s_n = \beta_n^{(+)}$, (75) donne bien un polynôme de degré $n+1$ qui vérifie $N_n M_n y = -\gamma_n y$. En reprenant (76) en intervertissant α_n et β_n , on trouve $N_n M_n y = p^2 y'' + p(p' + \alpha_n + \beta_n) y' + (p\alpha'_n + \alpha_n \beta_n) y = p(L - \lambda_n + \alpha'_n - \beta'_n) y - \gamma_n y$, ce qui correspond à une nouvelle valeur propre $\lambda_n - \alpha'_n + \beta'_n = \lambda_n - q' + p'' + 2\beta'_n = \lambda_{n+1}$ quand on prend $\beta_n^{(+)}$. On trouve ainsi des polynômes de tous les degrés. \square

Pour plus de détails, en particulier pour la **récurrence**, on examine la *deuxième* solution $\beta_n^{(-)}$. On obtient cette fois λ_{n-1} , d'où $\Phi_{n-1} = \text{constante}$ ($p\Phi'_n + \beta_n^{(-)}\Phi_n$), d'où la récurrence en éliminant $p\Phi'_n$ avec (75).

Un problème électrostatique (Stieltjes). Soit $n+2$ points $x_0 = 1 > x_1 > \dots > x_n > x_{n+1} = -1$ identiquement chargés, et soumis à une force répulsive inversement proportionnelle à la distance. Que valent

x_1, \dots, x_n à l'équilibre? Chaque x_i est soumis à une force $\sum_{\substack{j=0 \\ j \neq i}}^{n+1} \frac{1}{x_i - x_j} = 0, i = 1, \dots, n$. Soit $\Phi(x) = (x -$

$x_1) \dots (x - x_n)$ et $\Psi(x) = (x^2 - 1)\Phi(x)$. Comme $\frac{\Psi'(x)}{\Psi(x)} = \sum_{j=0}^{n+1} \frac{1}{x - x_j}$, on a $\lim_{x \rightarrow x_i} \frac{\Psi'(x)}{\Psi(x)} - \frac{1}{x - x_i} = \frac{\Psi''(x_i)}{2\Psi'(x_i)} = 0$,

pour $i = 1, \dots, n$. Le polynôme Ψ'' de degré n est nul en x_1, \dots, x_n , donc $\Psi''(x) = \text{const.} \Phi(x)$, ou $(x^2 - 1)\Phi''(x) + 4x\Phi'(x) = \text{const.} \Phi(x)$, ce qui correspond au polynôme de Jacobi $P_n^{(1,1)}$ (points de **Lobatto**).

On a imaginé plusieurs critères souvent très élégants²¹ de caractérisation des polynômes orthogonaux classiques:

- (1) Comme on vient de le voir, ce sont les seuls figurant entièrement dans des fonctions propres d'opérateurs de Sturm-Liouville.
- (2) Les dérivées de ces polynômes forment encore un système de polynômes orthogonaux (Sonin, Hahn).
- (3) Rodrigues (Hildebrandt)
- (4) Ils sont les seuls à vérifier une relation différentielle de type (75).

On peut étendre la formule de Rodrigues à tous les cas: **Théorème.** *La dérivée $n^{\text{ème}}$ d'une fonction n fois continûment dérivable R_n vérifiant*

$$R_n(a) = R'_n(a) = \dots = R_n^{(n-1)}(a) = R_n(b) = R'_n(b) = \dots = R_n^{(n-1)}(b) = 0$$

est orthogonale à \mathcal{P}_{n-1} selon le produit scalaire $(f, g)_1 = \int_a^b f(x)g(x)dx$. Il s'ensuit que si $R_n^{(n)}$ est le produit d'une fonction de poids w et d'un polynôme de degré n , ce dernier polynôme est le polynôme Φ_n orthogonal de degré n selon le produit scalaire $(f, g) = \int_a^b f(x)g(x)w(x)dx$.

²⁰Théorie des surfaces, II, Gauthier-Villars, 1915, articles 408–409 (= pages 210-216); on lira aussi, de F. Alberto Grünbaum et L. Haine: Orthogonal polynomials satisfying differential equations: the role of the Darboux transformation, *Centre de Recherches Mathématiques (CRM) Proceedings & Lecture Notes* 9 (1996) 143-154.

²¹cf. Al-Salam, *op. cit.*

En effet, soit $p \in \mathcal{P}_{n-1}$. Par intégrations par parties,

$$\int_a^b R_n^{(n)}(x)p(x)dx = - \int_a^b R_n^{(n-1)}(x)p'(x)dx = \dots = (-1)^n \int_a^b R_n(x)p^{(n)}(x)dx = 0,$$

les termes aux limites étant nuls.

Si, de plus, $R_n^{(n)}$ s'avère être de la forme $w\Phi_n$, où Φ_n est un polynôme de degré n , on a bien

$$\int_a^b \Phi_n(x)p(x) w(x) dx = 0, \quad \forall p \in \mathcal{P}_{n-1},$$

et Φ_n est bien le polynôme orthogonal de degré n par rapport à w . \square

La difficulté est évidemment de trouver une telle fonction R_n pour une densité w donnée. . . En principe, il “suffit” de résoudre l'équation différentielle linéaire d'ordre $2n$: $(R_n^{(n)}/w)^{(n)} = \text{constante}$ sous les $2n$ conditions aux limites $R_n^{(k)}(a) = R_n^{(k)}(b) = 0, k = 0, \dots, n-1$. Le recours à la fonction de Rodrigues R_n ne s'est révélé utile que dans le cas des polynômes orthogonaux classiques, où on peut montrer que l'on a $R_n(x) = \text{constante } w(x)(p(x))^n$, p étant le polynôme de degré ≤ 2 intervenant dans la définition de ces polynômes.

Legendre: $R_n(x) = (1-x^2)^n$, Tchebycheff 1^{ère} et 2^{ème} espèce $R_n(x) = (1-x^2)^{n\mp 1/2}$, Laguerre: $R_n(x) = x^{n+\alpha} \exp(-x)$, Hermite: $R_n(x) = \exp(-x^2)$.

3.11. Usages et variétés de polynômes orthogonaux.

Le recours de plus en plus fréquent au traitement numérique des problèmes les plus divers se traduit par une extension de l'usage des polynômes orthogonaux. On les rencontre en probabilités, en mécanique céleste, en physique de l'état solide, en mécanique des fluides (de l'ionosphère aux matières plastiques), en traitement du signal (compression d'images, reconnaissance de la parole, tomographie, radar), dans les théories les plus pointues de la gravité quantique, en mathématiques discrètes (codage, cryptographie), et même en théorie des nombres (fractions continues, nombres irrationnels et transcendants).

Les polynômes orthogonaux tiennent une place essentielle dans la boîte à outils de l'analyse numérique, pouvant servir à l'interpolation, au lissage, aux quadratures, comme on le verra encore plus loin.

Historiquement, on étudia d'abord les polynômes orthogonaux en relation avec des formules de quadrature (Gauss, Jacobi, Christoffel), des développements de potentiels en coordonnées sphériques (Legendre), des formules d'accélération de convergence (Laguerre, Stieltjes), et autres problèmes d'approximation (Tchebycheff, Hermite)²². On retrouve les polynômes de Laguerre en mécanique quantique (potentiels coulombiens), ainsi que les polynômes d'Hermite (oscillateur harmonique).

Les polynômes orthogonaux sur un domaine du plan complexe $\int_D \overline{\Phi_n(z)}\Phi_m(z) w(z) dx dy = 0, n \neq m$ portent le nom de polynômes de Carleman-Bergman. Ils sont utilisés en constructions de représentations conformes de domaines. Ils ne vérifient pas de relation de récurrence simple.

On a cependant le

Théorème (Fejér). *Soit μ une mesure positive sur \mathbb{C} de support $S = \{z : \mu(v) > 0 \text{ pour tout voisinage } v \text{ de } z\}$. Alors, les zéros du polynôme Φ_n orthogonal à \mathcal{P}_{n-1} par rapport à μ sont tous situés dans la fermeture convexe de S .*

²²cf. C. Brezinski, *History of Continued Fractions and Padé Approximants*, Springer-Verlag, 1991; J.A. Shohat, E. Hille, J.L. Walsh, A bibliography on orthogonal polynomials, *Bull. National Research Council*, n° 103, (August 1940), Nat. Acad. Sciences, Washington.

Démonstration intuitive: soit z_i un zéro de Φ_n . Φ_n est orthogonal à $\frac{\Phi_n(z)}{z - z_i} \in \mathcal{P}_{n-1}$:

$$0 = \int_S \Phi_n(z) \overline{\left(\frac{\Phi_n(z)}{z - z_i}\right)} d\mu(z) = \int_S (z - z_i) \left|\frac{\Phi_n(z)}{z - z_i}\right|^2 d\mu(z),$$

ou

$$z_i = \frac{\int_S z d\nu(z)}{\int_S d\nu(z)},$$

où ν est la mesure positive $d\nu(z) = \left|\frac{\Phi_n(z)}{z - z_i}\right|^2 d\mu(z)$. Le zéro z_i est donc le centre de gravité de S muni de la densité $d\nu$...

Les polynômes orthogonaux sur un arc ou un contour du plan complexe $\int_C \Phi_n(z) \overline{\Phi_m(z)} w(z) ds = 0, n \neq m$ sont appelés polynômes orthogonaux de Szegő. La théorie des polynômes orthogonaux sur un *cercle* est très riche, ces polynômes sont très utilisés en théorie du signal²³. La *récurrence* des polynômes orthogonaux sur le cercle unité a cette forme remarquable:

$$\Phi_{n+1}(z) = z\Phi_n(z) + \Phi_{n+1}(0) z^n \overline{\Phi_n(z^{-1})}.$$

Comme exemple de système connu, on a, pour les polynômes orthogonaux de Jacobi sur le cercle unité, $z = e^{i\theta}$, $w(z) = (1 - \cos \theta)^\alpha (1 + \cos \theta)^\beta$, $\Phi_{n+1}(0) = -[\alpha + (-1)^{n+1}\beta]/(n + \alpha + \beta + 1)$ (V.M. Badkov, Systems of orthogonal polynomials explicitly represented by the Jacobi polynomials, *Mat. Zametki* **42** (1987) 650–660 = *Math. Notes of the Academy of Sciences of the USSR, a translation of Mat. Zametki* **42** (1988) 858–863).

Revenons sur un intervalle, les polynômes orthogonaux relativement à un produit scalaire de Sobolev $(f, g) = \int_a^b [f(x)g(x)d\mu_1(x) + f'(x)g'(x)d\mu_2(x)]$ sont parfois utilisés dans le traitement numérique d'équations différentielles et aux dérivées partielles. Ils ne semblent en général pas vérifier de récurrence simple.

On rassemble ici quelque données sur les polynômes orthogonaux les plus courants,... et quelques autres:

²³ W.B. Jones, O. Njåstad, W.J. Thron, Moment theory, orthogonal polynomials, quadrature, and continued fractions associated with the unit circle, *Bull. London Math. Soc.* **21** (1989) 113–152; P. Delsarte, Y. Genin, On the role of orthogonal polynomials on the unit circle in digital signal processing applications, pp. 115–133 in *Orthogonal Polynomials: Theory and Practice*, (P. Nevai, ed.), NATO ASI Series C: Math. and Phys. Sci. **294**, Kluwer, 1990; B. Simon: *Orthogonal Polynomials on the Unit Circle: Part 1: Classical Theory; Part 2: Spectral Theory*, American Mathematical Society Colloquium Publications, Vol. **54** (2 vol.), 2005; M. E. H. Ismail: *Classical and Quantum Orthogonal Polynomials in One Variable*, Cambridge University Press, Encyc. of Mathematics and its Applications (No. 98), 2005.

α_n	β_n^2	Symbole	w ou μ	Nom	Usage
0	$\beta_1^2 = 1/2$ $1/4, n > 1$	T_n	$w(x) = (1 - x^2)^{-1/2},$ $-1 < x < 1$	Tchebycheff 1 ^{ère} espèce	analyse numérique
0	$1/4$	U_n	$w(x) = (1 - x^2)^{1/2},$ $-1 < x < 1$	Tchebycheff 2 ^{ème} espèce	analyse numérique
0	$\frac{n^2}{4n^2 - 1}$	P_n	$w(x) = 1,$ $-1 < x < 1$	Legendre	analyse sur la sphère et quadratures de Gauss
0	$\frac{n(n + 2\lambda - 1)}{4(n + \lambda)(n + \lambda - 1)}$ $\beta_1^2 = \frac{1}{2(\lambda + 1)}$	$C_n^\lambda =$ $P_n^{(\lambda-1/2, \lambda-1/2)}$	$w(x) = (1 - x^2)^{\lambda-1/2},$ $-1 < x < 1$ $\lambda > -1/2$	Gegenbauer, ou ultrasphériques	analyse sur la sphère si $\lambda - 1/2$ entier
*	*	$P_n^{(\alpha, \beta)}$	$w(x) = (1 - x)^\alpha(1 + x)^\beta,$ $-1 < x < 1$ $\alpha > -1, \beta > -1$	Jacobi	généralisation des précédents
$2n + \alpha + 1$	$n(n + \alpha)$	L_n^α	$w(x) = x^\alpha \exp(-x),$ $x > 0, \alpha > -1$	Laguerre	potentiels coulombiens, etc.
0	$n/2$	H_n	$w(x) = \exp(-x^2),$ $x > 0$	Hermite	oscillateur harmonique, etc.
$(a + b)/2$	$\frac{n^2(N^2 - n^2)h^2}{4(4n^2 - 1)}$ $h = \frac{b - a}{N - 1}$	t_n	μ constant par morceaux, avec même accroissement aux points $a + kh,$ $k = 0, 1, \dots, N - 1.$	Gram- Tchebycheff	moindres carrés

* Pour Jacobi, $\alpha_n = \frac{\beta^2 - \alpha^2}{(2n + \alpha + \beta)(2n + \alpha + \beta + 2)}, \beta_n^2 = \frac{4n(n + \alpha)(n + \beta)(n + \alpha + \beta)}{(2n + \alpha + \beta - 1)(2n + \alpha + \beta)^2(2n + \alpha + \beta + 1)}$

On appelle aussi polynômes de Tchebycheff de troisième et quatrième espèces les polynômes orthogonaux relatifs aux densités $(1 - x)^{1/2}(1 + x)^{-1/2}$ et $(1 - x)^{-1/2}(1 + x)^{1/2}$ sur $(-1, 1)$.

Voir aussi A. Erdélyi, W. Magnus, F. Oberhettinger, F.G. Tricomi, *Higher Transcendental Functions* (Bateman Manuscript Project), vol. 2, McGraw-Hill, 1953; M. Abramowitz, I.A. Stegun, *Handbook of Mathematical Functions*, National Bureau of Standards, Washington, 1964, = Dover, New York, 1965, etc.; T.S. Chihara, *An Introduction to Orthogonal Polynomials*, Gordon & Breach, 1978; A. Nikiforov, V. Ouvarov,

Fonctions spéciales de la physique mathématique, Mir, 1983, R. Koekoek, R.F. Swarttouw, The Askey-scheme of hypergeometric orthogonal polynomials and its q -analogue, Delft University of Technology, Faculty of Information Technology and Systems, Department of Technical Mathematics and Informatics, Report no. 98-17, 1998. See also the online version in <http://aw.twi.tudelft.nl/~koekoek/askey.html>

où on pourra tout découvrir sur les polynômes de Charlier, Meixner, Hahn, Krawtchouk, Lommel, Pollaczek, Stieltjes-Wigert, Tricomi-Carlitz, Heine, etc. etc.

Aspects stochastiques: W. Schoutens, *Stochastic Processes and Orthogonal Polynomials*, Springer Lect. Notes Stat. **146**, Springer-Verlag, New York, 2000.

3.12. Harmoniques sphériques et fonctions de Legendre.

Cf. H. Hochstadt, *Les fonctions de la physique mathématique* (trad. française S. Colombo), Masson, 1973, chap. 5 & 6; I.A. Stegun, chap. 9 de [Abr], chap. 8 de P. Frank & R. v. Mises, *Die Differential- und Integralgleichungen der Mechanik und Physik*, vol. 1, Vieweg, =Dover, 1961, etc.

Cours de Christophe Vigny, Laboratoire de géologie de l'École normale supérieure:

<http://www.geologie.ens.fr/~vigny/cours/chp-gphy-2.html>

L'opérateur de Laplace $\Delta := \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2 = \text{div grad}$ est formellement hermitien sur tout ensemble de fonctions suffisamment régulières s'annulant sur le bord d'un domaine D : par la formule de la divergence,

$$\int_D f \text{div grad } g \, dx dy dz = - \int_D \text{grad } f \cdot \text{grad } g \, dx dy dz$$

est bien symétrique en f et g (le terme aux limites $\int_{\partial D} f \frac{\partial g}{\partial n} \, dS$ étant nul).

Les fonctions propres seront donc orthogonales entre elles, selon le produit scalaire usuel des fonctions sur \mathbb{R}^3 .

En coordonnées sphériques $x = r \sin \theta \cos \varphi$, $y = r \sin \theta \sin \varphi$, $z = r \cos \theta$, le laplacien devient

$$\frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} + \frac{1}{r^2 \sin \theta} \frac{\partial(\sin \theta \frac{\partial}{\partial \theta})}{\partial \theta} + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \varphi^2} = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \Delta_B,$$

où $\Delta_B = \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \varphi^2}$ est l'opérateur de *Laplace-Beltrami* sur la sphère unité \mathcal{S} . On vérifie directement que Δ_B est formellement hermitien sur $\mathcal{C}^1(\mathcal{S})$:

$$\begin{aligned} \int_{\mathcal{S}} f \Delta_B g \, dS &= \int_0^\pi \sin \theta \, d\theta \int_0^{2\pi} d\varphi f \left(\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial g}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2 g}{\partial \varphi^2} \right) \\ &= \int_0^{2\pi} d\varphi \left\{ \left[f \sin \theta \frac{\partial g}{\partial \theta} \right]_{\theta=0}^{\theta=\pi} - \int_0^\pi \sin \theta \frac{\partial f}{\partial \theta} \frac{\partial g}{\partial \theta} \, d\theta \right\} + \int_0^\pi d\theta \left\{ \left[\frac{f}{\sin \theta} \frac{\partial g}{\partial \varphi} \right]_{\varphi=0}^{\varphi=2\pi} - \frac{1}{\sin \theta} \int_0^{2\pi} \frac{\partial f}{\partial \varphi} \frac{\partial g}{\partial \varphi} \, d\varphi \right\} \\ &= - \int_{\mathcal{S}} \left(\frac{\partial f}{\partial \theta} \frac{\partial g}{\partial \theta} + \frac{1}{\sin^2 \theta} \frac{\partial f}{\partial \varphi} \frac{\partial g}{\partial \varphi} \right) \, dS \end{aligned}$$

est bien symétrique en f et en g . On peut aussi construire des fonctions de 3 variables s'annulant au bord d'une boule B de \mathbb{R}^3 : $F(r, \theta, \varphi) = \rho(r) f(\theta, \varphi)$, avec $\rho(R) = 0$, et constater que le laplacien tridimensionnel de

$$F \text{ est } \Delta F = r^{-2} (r^2 \rho')' f + r^{-2} \rho \Delta_B f, \text{ donc } 0 = \int_B (F \Delta G - G \Delta F) r^2 \, dr \, dS = \int_0^R \rho^2 \, dr \int_{\mathcal{S}} (f \Delta_B g - g \Delta_B f) \, dS.$$

Une suite orthogonale de fonctions sur \mathcal{S} pourra donc être construite avec des fonctions propres de Δ_B .

On obtient de telles fonctions propres à partir de fonctions harmoniques (solutions de l'équation de Laplace $\Delta F = 0$) dans l'espace, de la forme $F(x, y, z) = r^n f(\theta, \varphi)$. En effet,

$$\Delta F = \Delta(r^n f) = 0 \Rightarrow \frac{\partial}{\partial r} \left(r^2 \frac{\partial r^n f}{\partial r} \right) + \Delta_B(r^n f) = 0 \Rightarrow \Delta_B f = -n(n+1)f.$$

La fonction harmonique fondamentale dans \mathbb{R}^3 est l'inverse de la distance à un point fixe P_0 , on la rencontre comme potentiel classique (gravitationnel ou électrostatique ou magnétostatique) dû à une masse ou charge

placée en P_0 . Soit $\rho(P_0, P)$ la distance entre P_0 et le point courant P . En coordonnées sphériques $r_0, \theta_0, \varphi_0, r, \theta, \varphi$,

$$\frac{1}{\rho(P_0, P)} = \frac{1}{\sqrt{r_0^2 - 2r_0r \cos \gamma + r^2}},$$

où γ est l'angle entre OP_0 et OP . On a $\cos \gamma = \cos \theta \cos \theta_0 + \sin \theta \sin \theta_0 \cos(\varphi - \varphi_0)$.

Soit $r_0 > r$ et développons $1/\rho$ en série de Taylor de r : $\rho^{-1} = r_0^{-1}(1 - 2(r/r_0) \cos \gamma + (r/r_0)^2)^{-1/2} = r_0^{-1} + r_0^{-2} \cos \gamma r + r_0^{-3}(3 \cos^2 \gamma - 1)r^2 + \dots$

On écrit $\rho^{-1} = (r_0^2 - 2r_0r \cos \gamma + r^2)^{-1/2} = \sum_{m=0}^{\infty} P_m(\cos \gamma) \frac{r^m}{r_0^{m+1}}$, $r < r_0$, (si $r > r_0$, on permute r et r_0 ²⁴), où P_m est le **polynôme de Legendre** de degré m .

Pour se convaincre que P_n est bien un polynôme de degré n , on retrouve la *réurrence*:

$$\partial \rho^{-1} / \partial r = (r_0 \cos \gamma - r)(r_0^2 - 2r_0r \cos \gamma + r^2)^{-3/2} = \sum_{m=0}^{\infty} (m+1)P_{m+1}(\cos \gamma) \frac{r^m}{r_0^{m+2}},$$

multiplions par $r_0^2 - 2r_0r \cos \gamma + r^2$:

$$\sum_{m=0}^{\infty} (m+1)P_{m+1}(\cos \gamma)(r_0^2 - 2r_0r \cos \gamma + r^2) \frac{r^m}{r_0^{m+2}} = \sum_{n=0}^{\infty} [(n+1)P_{n+1}(\cos \gamma) - 2n \cos \gamma P_n(\cos \gamma) + (n-1)P_{n-1}(\cos \gamma)] \frac{r^n}{r_0^n},$$

qui vaut également $(r_0 \cos \gamma - r)(r_0^2 - 2r_0r \cos \gamma + r^2)^{-1/2} = (r_0 \cos \gamma - r) \sum_{m=0}^{\infty} P_m(\cos \gamma) \frac{r^m}{r_0^{m+1}}$, d'où, en identifiant en $(r/r_0)^n$: $(n+1)P_{n+1}(\cos \gamma) - 2n \cos \gamma P_n(\cos \gamma) + (n-1)P_{n-1}(\cos \gamma) = \cos \gamma P_n(\cos \gamma) - P_{n-1}(\cos \gamma)$:

$$(n+1)P_{n+1}(x) = (2n+1)x P_n(x) - nP_{n-1}(x). \quad (77)$$

Remarquons que $P_n(1) = 1$, et que $P_n(x) = \frac{(2n-1)!!}{n!} x^n + \dots$.

Voyons maintenant que chaque $P_n(\cos \gamma)$ est une fonction propre de Δ_B :

$$0 = \Delta \rho^{-1} = \sum_0^{\infty} r_0^{-n-1} \Delta(r^n P_n(\cos \gamma)) = \sum_0^{\infty} r_0^{-n-1} r^{n-2} (n(n+1)P_n + \Delta_B P_n)$$

est le développement de Taylor de la fonction nulle, donc

$$\Delta_B P_n(\cos \gamma) = -n(n+1)P_n(\cos \gamma), \quad \cos \gamma = \cos \theta \cos \theta_0 + \sin \theta \sin \theta_0 \cos(\varphi - \varphi_0).$$

Equation différentielle des polynômes de Legendre: $\theta_0 = 0 \Rightarrow$

$$\frac{1}{\sin \theta} \frac{d}{d\theta} \left(\sin \theta \frac{dP_n(\cos \theta)}{d\theta} \right) + n(n+1)P_n(\cos \theta) = 0,$$

ou

$$\frac{d}{dx} \left((1-x^2) \frac{dP_n(x)}{dx} \right) + n(n+1)P_n(x) = (1-x^2) \frac{d^2 P_n(x)}{dx^2} - 2x \frac{dP_n(x)}{dx} + n(n+1)P_n(x) = 0 \quad (78)$$

Séparons (θ, φ) de (θ_0, φ_0) : $P_n(\cos \gamma) = P_n(\cos \theta \cos \theta_0 + \sin \theta \sin \theta_0 \cos(\varphi - \varphi_0))$ est une combinaison de produits de puissances $\cos^p \theta \cos^q \theta_0 \sin^q \theta \sin^q \theta_0 \cos^q(\varphi - \varphi_0)$, avec $p+q \leq n$. Regroupons les contributions

de φ en *série de Fourier* $P_n(\cos \gamma) = 2 \sum_{m=0}^n C_{n,m}(\theta, \theta_0) \cos(m(\varphi - \varphi_0))$.

On a $C_{0,0} = 1$; $C_{1,0}(\theta, \theta_0) = \cos \theta \cos \theta_0$, $C_{1,1}(\theta, \theta_0) = \sin \theta \sin \theta_0 / 2$, etc. Les $C_{n,m}(\theta, \theta_0)$ sont symétriques en θ et en θ_0 .

De la réurrence (77) des polynômes de Legendre, avec $x = \cos \gamma = P_n(\cos \theta \cos \theta_0 + \sin \theta \sin \theta_0 \cos(\varphi - \varphi_0))$, on tire les relations pour les $C_{n,m}$

$$C_{n+1,m} = \frac{2n+1}{n+1} \left[\cos \theta \cos \theta_0 C_{n,m} + \sin \theta \sin \theta_0 \frac{C_{n,m-1} + C_{n,m+1}}{2} \right] - \frac{n}{n+1} C_{n-1,m},$$

²⁴On a donc alors une série de puissances *néglatives* de r , convergente dans l'extérieur de la boule de rayon r_0 .

d'où on peut montrer que $C_{n,m}(\theta, \theta_0) = \sin^m \theta \sin^m \theta_0 D_{n,m}(\theta, \theta_0)$, où $D_{n,m}(\theta, \theta_0)$ est un polynôme de degré $n - m$ en $\cos(\theta$ et $\cos \theta_0$. En particulier, $\theta_0 = 0 \Rightarrow P_n(\cos \theta) = D_{n,0}(\theta, 0)$. Enfin, on porte

$$P_n(\cos \gamma) = \sum_{m=0}^n \sin^m \theta \sin^m \theta_0 D_{n,m}(\theta, \theta_0) \cos(m(\varphi - \varphi_0))$$

dans $\Delta_B P_n(\cos \gamma) = -n(n+1)P_n(\cos \gamma)$:

$$\sum_{m=0}^n \sin^m \theta_0 \left\{ \frac{1}{\sin \theta} \frac{d}{d\theta} \left[\sin \theta \frac{d}{d\theta} (\sin^m \theta D_{n,m}) \right] + \left(n(n+1) - \frac{m^2}{\sin^2 \theta} \right) \sin^m \theta D_{n,m} \right\} \cos(m(\varphi - \varphi_0)) = 0,$$

ce qui donne $\frac{d^2}{d\theta^2} D_{n,m} + (2m+1) \cotg \theta \frac{d}{d\theta} D_{n,m} + (n(n+1) - m(m+1)) D_{n,m} = 0$,

ou, en $x = \cos \theta$, $(1-x^2) \frac{d^2}{dx^2} D_{n,m} - 2(m+1) \frac{d}{dx} D_{n,m} + (n(n+1) - m(m+1)) D_{n,m} = 0$. On reconnaît l'équation différentielle des dérivées $m^{\text{èmes}}$ de (78), et les seules solutions polynomiales sont $D_{n,m}(\theta, \theta_0) = P_n^{(m)}(\cos \theta)$ fois une fonction de θ_0 , soit $K_{n,m} P_n^{(m)}(\cos \theta) P_n^{(m)}(\cos \theta_0)$ pour respecter la symétrie en θ et θ_0 .

La récurrence des $C_{n,m}$, donc des $D_{n,m}$, avec $\theta_0 = 0$, devient

$$K_{n+1,m} P_{n+1}^{(m)}(x) P_{n+1}^{(m)}(1) = \frac{2n+1}{n+1} K_{n,m} x P_n^{(m)}(x) P_n^{(m)}(1) + \text{degré} \leq n,$$

ce qui aboutit à $K_{n,m} = \frac{(n-m)!}{(n+m)!}$.

On note $P_n^m(x) = (1-x^2)^{m/2} \frac{d^m}{dx^m} P_n(x)$, on a donc

$$P_n(\cos \gamma) = P_n(\cos \theta \cos \theta_0 + \sin \theta \sin \theta_0 \cos(\varphi - \varphi_0)) = 2 \sum_{m=0}^n \frac{(n-m)!}{(n+m)!} P_n^m(\cos \theta) P_n^m(\cos \theta_0) \cos(m(\varphi - \varphi_0)).$$

Les fonctions $Y_{n,m}(\theta, \varphi) = \frac{1}{\sqrt{2\pi}} (\cos m\varphi \text{ et } \sin m\varphi) \sqrt{\frac{2n+1}{2} \frac{(n-m)!}{(n+m)!}} P_{n,m}(\cos \theta)$ sont appelées

harmoniques sphériques, elles sont orthonormales sur la sphère.

On a bien $\int_{-1}^1 P_{n_1}^m(x) P_{n_2}^m(x) dx = 0$ si $n_1 \neq n_2$, et comme $P_n^m(x) = \sin^m \theta$ fois un polynôme de degré $n - m$ en $x = \cos \theta$, ces polynômes, pour m fixé, sont orthogonaux par rapport à la fonction de poids $\sin^{2m} \theta = (1-x^2)^m$.

Pour un essai intéressant de O. de Viron sur le caractère total de la suite des harmoniques sphériques, s'adresser à l'auteur:

```
*****
* Olivier de Viron *
* Royal Observatory of Belgium Tel: 32-2-3730312 *
* Observatoire Royal de Belgique Fax: 32-2-3749822 *
* 3, avenue Circulaire e-mail: deviron@oma.be *
* B-1180 Bruxelles, Belgium *
*****
```

cf. aussi

sci.math.num-analysis #42078 (3 + 345 more)

[1]

From: fractalier@aol.com (Fractalier)

[1] Spherical Harmonics Visualization Study

Date: Wed May 06 16:37:32 EDT 1998

Dear Mathematicians:

I have recently uploaded over 400 three-dimensional visualizations of various

spherical harmonics with data for your research and enjoyment. They can be found at:

<http://www.lifesmith.com/spharmin.html>

Thank you for your support.

Jeff Berkowitz, M.S.

Lifesmith Classic Fractals

End of article 42078 (of 42246) -- what next? [npq]

3.13. Polynômes d’Hermite et mécanique quantique.

Les polynômes d’Hermite

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2} = 2^n x^n - 2^{n-1} 1!! \binom{n}{2} x^{n-2} + 2^{n-2} 3!! \binom{n}{4} x^{n-4} - \dots$$

sont orthogonaux sur \mathbb{R} par rapport à la fonction de poids e^{-x^2} :

$$\int_{-\infty}^{\infty} H_n(x) H_m(x) e^{-x^2} dx = \int_{-\infty}^{\infty} \frac{d^n}{dx^n} (H_m(x) e^{-x^2}) dx = n! 2^n \sqrt{\pi} \delta_{n,m}$$

(prendre $m \leq n$, sinon permuter n et m).

On a $H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x)$ et $H'_n(x) = 2nH_{n-1}(x)$.

Les fonctions du cylindre parabolique ou fonctions de Weber-Hermite

$$D_n(x) = 2^{-n/2} e^{-x^2/4} H_n\left(\frac{x}{\sqrt{2}}\right)$$

sont donc orthogonales dans $L^2(\mathbb{R})$. Les fonctions orthonormales sont $D_n(x)/\sqrt{n!\sqrt{2\pi}}$.

De la récurrence $D_{n+1}(x) = xD_n(x) - nD_{n-1}(x)$, donc,

$$x \frac{D_n(x)}{\sqrt{n!\sqrt{2\pi}}} = \sqrt{n+1} \frac{D_{n+1}(x)}{\sqrt{(n+1)!\sqrt{2\pi}}} + \sqrt{n} \frac{D_{n-1}(x)}{\sqrt{(n-1)!\sqrt{2\pi}}},$$

on voit que l’opérateur \mathbf{q} de multiplication par x agit sur les combinaisons des $D_n(x)/\sqrt{n!\sqrt{2\pi}}$ au moyen de la

matrice
$$\mathbf{q} = \begin{bmatrix} 0 & 1 & & & \\ 1 & 0 & \sqrt{2} & & \\ & \sqrt{2} & 0 & \sqrt{3} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}.$$

On a aussi $2D'_n(x) = -D_{n+1}(x) + (2n-1)D_{n-1}(x)$,

d’où la représentation de l’opérateur de dérivation
$$\frac{d}{dx} = \frac{1}{2} \begin{bmatrix} 0 & -1 & & & \\ 1 & 0 & -\sqrt{2} & & \\ & \sqrt{2} & 0 & -\sqrt{3} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}.$$
 On préfère garder

un opérateur hermitien en prenant plutôt

$$\mathbf{p} = \frac{d}{idx} = \frac{1}{2} \begin{bmatrix} 0 & i & & & \\ -i & 0 & i\sqrt{2} & & \\ & -i\sqrt{2} & 0 & i\sqrt{3} & \\ & & \ddots & \ddots & \ddots \end{bmatrix}.$$

Et on constate $\mathbf{pq} - \mathbf{qp} = i$ fois la matrice unité! Relation de base de la mécanique quantique. Toute fonction d’onde peut en effet s’exprimer comme une combinaison des fonctions D_n .

Les D_n sont d’ailleurs les fonctions d’onde d’un système particulier, l’oscillateur linéaire. La matrice $\mathbf{p}^2 + \mathbf{q}^2/4$ est diagonale d’éléments diagonaux $n + 1/2$, $n = 0, 1, \dots$

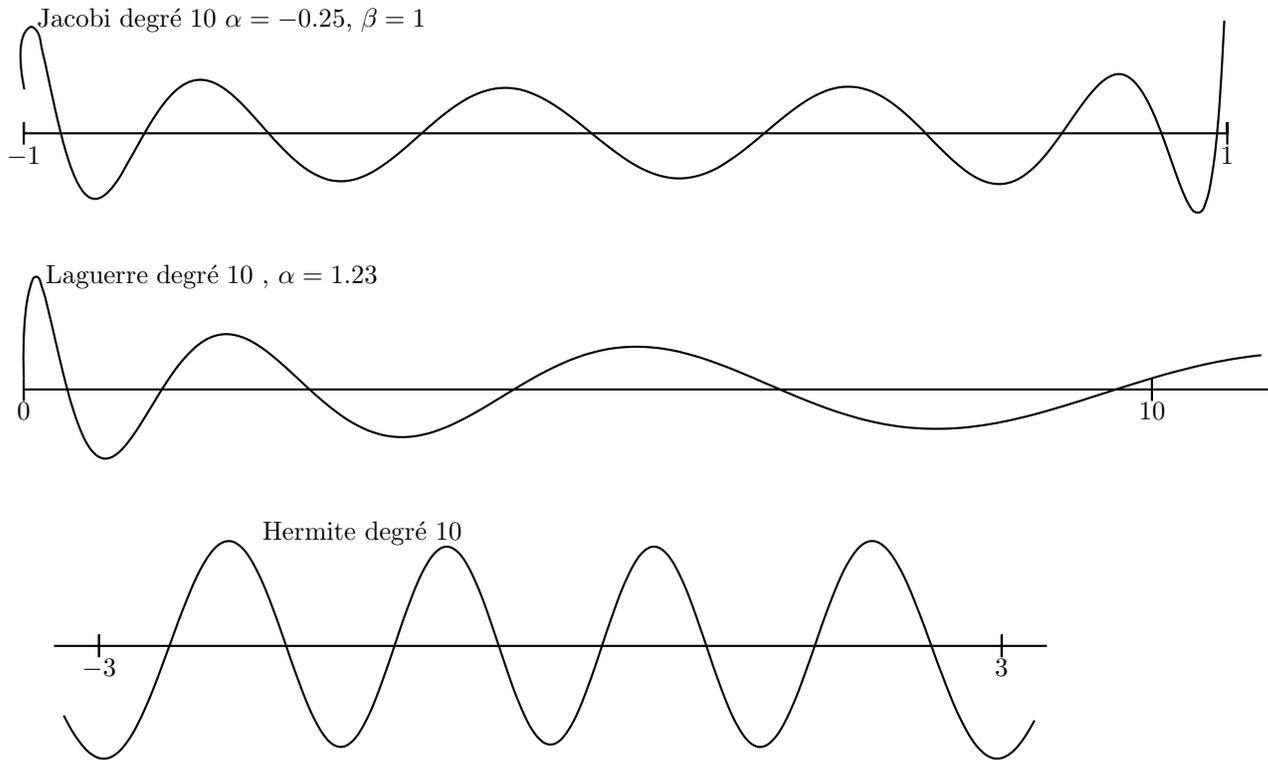
En variables physiques²⁵: $\psi_n(x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} \frac{1}{\sqrt{2^n n!}} e^{-\frac{m\omega}{2\hbar}x^2} H_n\left(x\sqrt{\frac{m\omega}{\hbar}}\right)$. Niveaux d'énergie: $E = (n + 1/2)\hbar\omega$, $n = 0, 1, \dots$. L'énergie varie donc par pas de $h = 2\pi\hbar$ fois la fréquence $\nu = \omega/(2\pi)$.
 $h = 6.626196 \cdot 10^{-34}$ Joule seconde.

Pour une controverse intéressante, voir B.L. van der Waerden: From matrix mechanics and wave mechanics to unified quantum mechanics, *Notices of the AMS* **44** (1997) 323-328.

3.14. Orthogonalité et équi-oscillation.

On a d'abord rencontré les polynômes de Tchebycheff à partir de conditions déqui-oscillation. Puis on a constaté que ces polynômes sont orthogonaux par rapport à la fonction de poids $(1-x^2)^{-1/2}$ sur $(-1, 1)$.

Si $\{\Phi_n\}$ est une suite de polynômes orthogonaux par rapport à une fonction de poids w sur $(-1, 1)$, $\{(1-x^2)^{1/4}w(x)^{1/2}\Phi_n(x)\}$ sont des fonctions orthogonales par rapport à $(1-x^2)^{-1/2}$, on constate un comportement raisonnablement équi-oscillant:



Polynômes de Jacobi, Laguerre et Hermite.

La figure ci-dessus montre les graphes de polynômes de Jacobi, Laguerre et Hermite, multipliés par $(1-x)^{\alpha/2+1/4}(1+x)^{\beta/2+1/4}$ dans le cas Jacobi, par $x^{\alpha/2+1/4}\exp(-x/2)$ dans le cas Laguerre par $\exp(-x^2/2)$ pour Hermite.

cf. `orthclas.m`

Théorème de Sonine-Pólya²⁶. Si Φ_n est un polynôme orthogonal classique sur $[a, b]$, et si $w(x)\sqrt{p(x)}$ est croissante (resp. décroissante) sur $[c, d] \subseteq [a, b]$, les ordonnées des maxima locaux de $|\Phi_n|$ sur $[c, d]$ forment une suite décroissante (resp. croissante).

²⁵Landau & Lifchitz, *Mécanique quantique*, § 23.

²⁶M. Redheffer, Monotonicity of the maxima, *American Math. Monthly* **105** (1998) 945-948.

En effet, les carrés de ces ordonnées sont les valeurs de $F_n(x) := \Phi_n^2(x) - \frac{p(x)}{\lambda_n} \Phi_n'^2(x)$ aux zéros de $p\Phi_n'$.
Dérivons F_n :

$$\begin{aligned} F_n' &= 2\Phi_n\Phi_n' - \frac{p'}{\lambda_n}\Phi_n'^2 - 2\frac{p}{\lambda_n}\Phi_n'\Phi_n'' \\ &= \Phi_n' \left[2\Phi_n - \frac{p'}{\lambda_n}\Phi_n' - 2\frac{p\Phi_n''}{\lambda_n} \right] \\ &= \Phi_n' \left[2\Phi_n - \frac{p'}{\lambda_n}\Phi_n' - 2\frac{\lambda_n\Phi_n - \rho\Phi_n'}{\lambda_n} \right] \\ &= \frac{2\rho - p'}{\lambda_n}\Phi_n'^2 \end{aligned}$$

où on a utilisé (72). D'où la thèse, puisque, sur un intervalle où $w\sqrt{p}$ est monotone, $(w\sqrt{p})'/(w\sqrt{p}) = w'/w + p'/(2p) = r/p + p'/(2p) = (2\rho - p')/(2p)$. (et $\lambda_n < 0$). \square

Un théorème de Szegő ([Sze] § 7.2). Si la densité w est croissante sur (c, b) avec $a \leq c < b$, $\sqrt{w}|\Phi_n|$ atteint son maximum sur $[c, b]$ en b .

En effet: soit $c \leq x \leq b$, $w(b)\Phi_n^2(b) - w(x)\Phi_n^2(x) = 2 \int_x^b \Phi_n(t)\Phi_n'(t)w(t)dt + \int_x^b \Phi_n^2(t)dw(t)$. 1.) Si x est entre le plus grand zéro de Φ_n et b , $\Phi_n(t)$ et $\Phi_n'(t)$ sont positifs, donc le résultat est positif. 2.) Si x est entre a et le plus petit zéro de Φ_n , $\int_x^b \Phi_n(t)\Phi_n'(t)w(t)dt = \underbrace{\int_a^b}_{=0} - \int_a^x \Phi_n(t)\Phi_n'(t)w(t)dt$ encore positif puisque Φ_n

et Φ_n' ont les signes $(-1)^n$ et $(-1)^{n-1}$ à gauche du plus petit zéro; 3.) Enfin, si x est entre deux zéros de Φ_n , on considère le polynôme $\Phi_n(t)$ divisé par le produit des $t - z_i$, où les z_i sont les zéros plus petits que x . Ce polynôme est orthogonal par rapport à la densité $w(t) \prod_{z_i < x} (t - z_i)^2$. On se ramène ainsi au cas 2.). \square

4. Moindres carrés, régression.

La construction d'approximations au sens des moindres carrés est une des applications numériques les plus utiles qui soient. A partir de principes établis dès le début du XIX^{ème} siècle par Legendre et Gauss, on a traité des problèmes de plus en plus considérables. L'élaboration de bons algorithmes se base sur les idées force de meilleure approximation en moyenne quadratique et d'utilisation de bases orthogonales.

Fixons le cadre:

On dispose d'observations y_0, y_1, \dots, y_N d'un phénomène $f(t)$ (t peut représenter ici une ou plusieurs variables indépendantes) que l'on soupçonne pouvoir être représenté par une forme

$$f(t) = \sum_{j=0}^n \alpha_j f_j(t),$$

où les fonctions f_j sont connues, mais où les coefficients α_j sont à déterminer ²⁷. Si cette hypothèse était vraie, et si chaque mesure y_i donnait exactement $f(t_i)$, on trouverait les α_j en extrayant n équations de

$$\mathbf{y} = \mathbf{f} = \mathbf{\Phi}\boldsymbol{\alpha},$$

²⁷Nombreux exemples en physique, économie, etc. etc. A noter que l'on se limite ici à des problèmes où les inconnues α_j interviennent linéairement, ce qui est une grosse simplification vis à vis des questions qui se posent le plus fréquemment.

où \mathbf{y} est le vecteur des y_i , Φ est la matrice des $f_j(t_i)$, $j = 0, \dots, n$, $i = 0, \dots, N \geq n$, α est le vecteur des α_j . En fait, les mesures n'étant pas exactes, on a plutôt

$$\mathbf{y} = \mathbf{f} + \mathbf{e} = \Phi\alpha + \mathbf{e},$$

on renonce à calculer exactement α , mais on va estimer ce vecteur par \mathbf{a} qui minimise la somme des carrés des résidus

$$\|\mathbf{y} - \Phi\mathbf{b}\|^2 = \sum_{i=0}^{i=N} \left(y_i - \sum_{j=0}^{j=n} f_j(t_i)b_j \right)^2$$

sur $\mathbf{b} \in \mathbb{R}^{n+1}$.

C'est bien un problème de meilleure approximation de $\mathbf{y} \in \mathbb{R}^{N+1}$ par un élément de V , espace vectoriel de dimension $n + 1$ sous-tendu par les colonnes de Φ , au sens de la norme Euclidienne de \mathbb{R}^{N+1} (on a donc supposé que les $n + 1$ colonnes de Φ sont des vecteurs indépendants de \mathbb{R}^{N+1}).

La solution $\Phi\mathbf{b}$ est donc la projection orthogonale de \mathbf{y} sur V . La manière traditionnelle de trouver \mathbf{a} consiste à écrire que le résidu $\mathbf{r} = \mathbf{y} - \Phi\mathbf{a}$ est orthogonal aux colonnes de Φ , c'est-à-dire, à former les **équations normales**:

$$\Phi^T \Phi \mathbf{a} = \Phi^T \mathbf{y}.$$

A noter que l'on retrouve bien la **matrice de Gram** de la base de V .

L'usage des équations normales n'est recommandé que pour de petites valeurs de n , le système étant mal conditionné si n est grand.

On sait maintenant l'intérêt de disposer de bases orthogonales. Si on a orthonormalisé les colonnes $\mathbf{f}_0, \dots, \mathbf{f}_n$ de Φ en $\mathbf{f}_0^\perp, \dots, \mathbf{f}_n^\perp$, il reste, par (58) (p. 79):

$$P\mathbf{y} = \Phi\mathbf{a} = \Phi^\perp \mathbf{a}^\perp = \sum_{j=0}^{j=n} (\mathbf{f}_j^{\perp T} \mathbf{y}) \mathbf{f}_j^\perp,$$

donc, $\mathbf{a}_j^\perp = \mathbf{f}_j^{\perp T} \mathbf{y}$, ou encore $\mathbf{a}^\perp = \Phi^{\perp T} \mathbf{y}$.

Si on a utilisé la méthode de Gram-Schmidt pour orthonormaliser les colonnes de Φ , le passage de Φ à Φ^\perp est $\Phi = \Phi^\perp \mathbf{R}$, où \mathbf{R} est une matrice triangulaire supérieure, $\mathbf{R}\mathbf{a} = \mathbf{a}^\perp$. Comme $\Phi^{\perp T} \Phi^\perp = \mathbf{I}$, $\mathbf{R}^T \mathbf{R} = \Phi^T \Phi$, factorisation de Cholesky de la matrice de Gram.

On obtient également \mathbf{R} en multipliant Φ à gauche par des matrices orthogonales appropriées (rotations de Givens ou réflexions de Householder) de façon à avoir $\Phi = \mathbf{Q}\tilde{\mathbf{R}}$, où $\tilde{\mathbf{R}}$ est une matrice rectangulaire de $N + 1$ lignes et $n + 1$ colonnes, la $j^{\text{ème}}$ colonne n'ayant que ses j premiers éléments non nuls (factorisation \mathbf{QR}). Alors, \mathbf{a}^\perp est le vecteur formé des $n + 1$ premiers éléments de $\mathbf{Q}^T \mathbf{y}$.

En matlab:

help polyfit

POLYFIT Polynomial curve fitting.

POLYFIT(x,y,n) finds the coefficients of a polynomial $p(x)$ of degree n that fits the data, $p(x(i)) \sim y(i)$, in a least-squares sense.

[p,S] = POLYFIT(x,y,n) returns the polynomial coefficients p and a matrix S for use with POLYVAL to produce error estimates on predictions.

If the errors in the data, y , are independent normal with constant variance, POLYVAL will produce error bounds which contain at least 50% of the predictions.

See also POLY, POLYVAL, ROOTS.

help slash

...
 If A is an M -by- N matrix with $M < \text{or} > N$ and B is a column vector with M components, or a matrix with several such columns, then $\mathbf{X} = A \backslash B$ is the solution in the least squares sense to the under- or overdetermined system of equations $A * X = B$. The effective rank, K , of A is determined from the QR decomposition with pivoting. A solution X is computed which has at most K nonzero components per column. If $K < N$ this will usually not be the same solution as $\text{PINV}(A) * B$. $A \backslash \text{EYE}(\text{SIZE}(A))$ produces a generalized inverse of A .

Voir aussi LSCOV.

Remarque. Si chaque ligne de Φ est formée de valeurs de fonctions en un même point: $[\varphi_0(t_i), \dots, \varphi_n(t_i)]$, et s'il y a exactement autant d'équations que d'inconnues, alors la solution

$$\sum_{j=0}^n a_j^\perp \varphi_j,$$

où $a_j^\perp = \varphi_j^{\perp T} \mathbf{y}$, ou encore $\mathbf{a}^\perp = \Phi^{\perp T} \mathbf{y}$,

interpole les données y_0, \dots, y_n aux points t_0, \dots, t_n , c'est-à-dire que l'on a

$$\sum_{j=0}^n a_j^\perp \varphi_j^\perp(t_i) = y_i, \quad i = 0, 1, \dots, n \quad (79)$$

En effet $\Phi^\perp \mathbf{a}^\perp = \Phi^\perp \Phi^{\perp T} \mathbf{y} = \mathbf{y}$, Φ^\perp étant une matrice carrée orthogonale.

Si les fonctions f_j sont des polynômes d'une variable réelle, on dispose de techniques particulières basées sur la relation de récurrence²⁸.

Voici une justification statistique de ces techniques:

Théorème de Gauss-Markoff²⁹. *Si les erreurs $e_i = y_i - f(t_i)$ sont des variables aléatoires indépendantes de moyenne nulle (estimation sans biais) et de même variance σ^2 , a_j obtenu par moindres carrés est l'estimateur linéaire sans biais de α_j de variance minimale. Le carré de la norme du résidu $\|\mathbf{r}\|^2 = \|\mathbf{y} - \Phi \mathbf{a}\|^2$ est alors une variable aléatoire de moyenne $(N - n)\sigma^2$.*

En effet, étudions un estimateur b_j de α_j ou, plus généralement, un estimateur b d'une combinaison linéaire $\beta = \sum_{j=0}^n \lambda_j \alpha_j$, linéaire en les y_i : $b = \sum_{i=0}^N \ell_i y_i = \boldsymbol{\ell}^T \mathbf{y}$. Exprimons d'abord que b est sans biais, c'est-à-dire que l'on a $\mathcal{E}(b) = \beta = \boldsymbol{\lambda}^T \boldsymbol{\alpha}$, $\forall \boldsymbol{\alpha}$: $\mathcal{E}(b) = \mathcal{E}(\boldsymbol{\ell}^T \mathbf{y}) = \boldsymbol{\ell}^T \mathcal{E}(\mathbf{f} + \mathbf{e}) = \boldsymbol{\ell}^T \mathbf{f} = \boldsymbol{\ell}^T \Phi \boldsymbol{\alpha}$, donc $\Phi^T \boldsymbol{\ell} = \boldsymbol{\lambda}$, résolu par $\boldsymbol{\ell} = \Phi \mathbf{x} + \mathbf{z}$, avec $\mathbf{x} = (\Phi^T \Phi)^{-1} \boldsymbol{\lambda}$ et $\mathbf{z} \perp V$.

Passons maintenant à la variance de b : $\mathcal{E}((b - \boldsymbol{\lambda}^T \boldsymbol{\alpha})^2) = \mathcal{E}((\boldsymbol{\ell}^T \mathbf{y} - \boldsymbol{\ell}^T \Phi \boldsymbol{\alpha})(\mathbf{y}^T \boldsymbol{\ell} - \boldsymbol{\alpha}^T \Phi^T \boldsymbol{\ell})) = \boldsymbol{\ell}^T \mathcal{E}((\mathbf{y} - \mathbf{f})(\mathbf{y} - \mathbf{f})^T) \boldsymbol{\ell} = \sigma^2 \boldsymbol{\ell}^T \boldsymbol{\ell} = \sigma^2 (\|\Phi \mathbf{x}\|^2 + \|\mathbf{z}\|^2)$ est donc minimal quand $\mathbf{z} = 0$, donc quand $b = \boldsymbol{\ell}^T \mathbf{y} = \boldsymbol{\lambda}^T (\Phi^T \Phi)^{-1} \Phi^T \mathbf{y} = \boldsymbol{\lambda}^T \mathbf{a}$, ce qui est bien l'estimateur par moindres carrés.

Enfin, $\mathcal{E}(\|\mathbf{r}\|^2) = \mathcal{E}((\mathbf{y} - \Phi \mathbf{a})^T (\mathbf{y} - \Phi \mathbf{a})) = \mathcal{E}((\mathbf{y} - \Phi (\Phi^T \Phi)^{-1} \Phi^T \mathbf{y})^T (\mathbf{y} - \Phi (\Phi^T \Phi)^{-1} \Phi^T \mathbf{y})) = \mathcal{E}((\mathbf{e} - \Phi (\Phi^T \Phi)^{-1} \Phi^T \mathbf{e})^T (\mathbf{e} - \Phi (\Phi^T \Phi)^{-1} \Phi^T \mathbf{e})) = \mathcal{E}(\mathbf{e}^T (\mathbf{I} - \Phi (\Phi^T \Phi)^{-1} \Phi^T)^2 \mathbf{e}) = \mathcal{E}(\mathbf{e}^T (\mathbf{I} - \Phi (\Phi^T \Phi)^{-1} \Phi^T) \mathbf{e}) = (N + 1)\sigma^2 - \sigma^2 \text{trace}(\Phi (\Phi^T \Phi)^{-1} \Phi^T) = (N - n)\sigma^2$, où on a utilisé $\mathbf{y} = \Phi \boldsymbol{\alpha} + \mathbf{e}$, $\mathcal{E}(e_i e_j) = \sigma^2 \delta_{i,j}$, $\mathcal{E}(\mathbf{e}^T \mathbf{X} \mathbf{e}) = \mathcal{E}(\sum_i \sum_j e_i X_{i,j} e_j) = \sum_i X_{i,i} \mathcal{E}(e_i^2) = \sigma^2 \text{trace}(\mathbf{X})$, et $\text{trace}(\mathbf{A} \mathbf{B}) = \text{trace}(\mathbf{B} \mathbf{A})$. \square

²⁸G.E. Forsythe, Generation and use of orthogonal polynomials for data-fitting with a digital computer, *J. Soc. Indust. Appl. Math.* **5** (1957) 74-88.

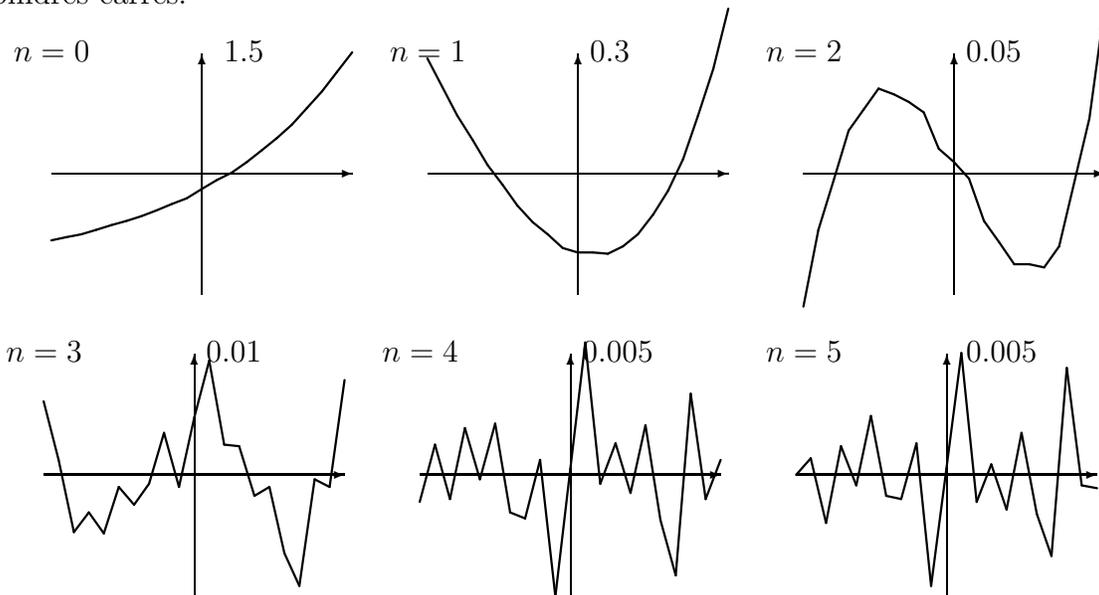
²⁹B.L. van der Waerden, *Mathematische Statistik*, Springer, 1957 = *Statistique mathématique*, Dunod, 1967, § 30-32.

Que se passe-t-il quand on ne sait pas exactement où il faut estimer f ? Réponse: on essaye des sous-espaces V de plus en plus grands (mais pas trop grands...), et on s'arrête quand $\mathbf{y} - \Phi\mathbf{a}$ manifeste un comportement raisonnablement aléatoire.

Comme expérience (pp. 266-267 de S.D. Conte et C. de Boor, *Elementary Numerical Analysis, An Algorithmic Approach*, 3rd ed., McGraw-Hill, 1981), on considère les 21 données³⁰

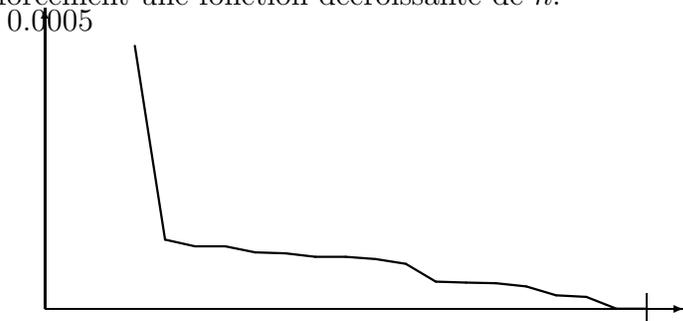
$x = -1.0 -0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0.00 0.10 0.20 0.30 0.40 0.50 0.60 0.70 0.80 0.90 1.00$
 $y = 0.37 0.41 0.45 0.50 0.55 0.61 0.67 0.74 0.82 0.90 1.00 1.11 1.22 1.35 1.49 1.65 1.82 2.01 2.23 2.46 2.72$

On examine $y(x) - p_n(x)$, où p_n est la meilleure approximation de degré $\leq n$ au sens des moindres carrés:



Tant que $n \leq 3$, l'erreur (qui a d'ailleurs un petit air de polynôme de Tchebycheff...) est systématique. A partir de $n = 4$, on a raisonnablement capturé le signal et on voit du bruit aléatoire. *Il ne sert à rien de dépasser $n = 4$...*

Voyons maintenant les sommes de carrés $\sum_{i=1}^N [y_i - p_n(x_i)]^2$ (le " χ^2 ") en fonction de n , forcément une fonction décroissante de n :



On voit bien le comportement en constante $\frac{1}{N-n}$ à partir de $n = 4$.

Cf. `cboor.m`

³⁰construites très artificiellement: chaque valeur de y est e^x arrondie à deux chiffres significatifs après le point décimal...

5. Approximation en norme $\| \cdot \|_1$.

On approche f par une combinaison linéaire p des fonctions Φ_0, \dots, Φ_n données sur $[a, b]$, et on cherche à minimiser

$$\|f - p\|_1 = \int_a^b |f(x) - p(x)| w(x) dx = \int_a^b S_p(x) (f(x) - p(x)) w(x) dx,$$

où w est une fonction poids donnée. La fonction S_p désigne le signe de $f - p$. Comparons deux approximations:

$$\|f - q\|_1 - \|f - p\|_1 = \int_a^b (S_q(x) - S_p(x)) (f(x) - q(x)) w(x) dx + \int_a^b S_p(x) (p(x) - q(x)) w(x) dx.$$

Si p et q sont proches, et si toutes les fonctions sont continues, la première intégrale est de l'ordre du carré de la deuxième. La condition de minimalité de $\|f - p\|_1$ est donc qu'il existe une fonction S , qui vaut partout 1 ou -1 , et qui soit orthogonale à Φ_0, \dots, Φ_n . Cette fonction S ne dépend que de w et n . Pour f donnée, $p = \hat{p}$ est alors la combinaison de Φ_0, \dots, Φ_n interpolant f aux points de discontinuité de S (*points canoniques*), si $f - p$ ne change de signe en aucun autre point de $[a, b]$.

Si $w = \text{constante}$ sur $[-1, 1]$ et $\overline{\{\Phi_0, \dots, \Phi_n\}} = \mathcal{P}_n$, alors les points canoniques sont les zéros du polynôme de Tchebycheff de deuxième espèce U_{n+1} .

Cf. [DeVLor, chap. 3, § 10].

6. Séries de Fourier en analyse numérique.

L'analyse harmonique consiste à extraire d'une fonction les harmoniques qu'elle contient en donnant à chacune le poids qui lui convient. La synthèse consiste à reproduire la fonction à partir de ses harmoniques.

J.P. Kahane, cité dans M. Willem, *Analyse harmonique réelle*, Hermann, Paris, 1995, p. 93.

Ce qu'on appelle maintenant les séries et intégrales de Fourier a une importance et une généralité qui dépassent de beaucoup leur but initial. Il y a d'innombrables domaines de la physique et de la technique qui font un usage des méthodes développées par Fourier. Tous les problèmes d'analyse d'images y font appel, par exemple ceux qui sont rencontrés dans les scanners à rayons X ou les échographes à ultrasons.

La recherche de la structure des grandes molécules au moyen de techniques cristallographiques, qui est à la base de la biologie moléculaire, fait abondamment usage des transformées de Fourier, et certains progrès récents dans ce domaine sont dus au fait que les calculatrices arrivent à effectuer de plus en plus rapidement ces transformations.

G. Charpak (Nobel physique 1992), cité dans J. Lacouture, *Champollion, une vie de lumières*, Grasset, 1988= Le Livre de Poche 6995, 1991.

L'analyse de Fourier a d'abord servi à quantifier le contenu fréquentiel de phénomènes ondulatoires. On a ensuite utilisé des séries de Fourier pour décrire des structures périodiques (cristaux). On traite maintenant de façon purement numérique des signaux (son, image) par techniques de Fourier. Ces méthodes se retrouvent dans des applications les plus diverses qui vont de la médecine à l'astrophysique.

On utilise en mathématiques les méthodes de Fourier dans l'étude d'opérateurs différentiels (initialement, Fourier s'est occupé de l'équation de la chaleur $\partial T/\partial t = K\partial^2 T/\partial x^2$ ³¹), la théorie des fonctions et de la mesure (des points fins de la théorie des ensembles et de la théorie de l'intégration sont apparus à partir de séries de Fourier), et on va jusqu'à trouver des séries de Fourier dans des méthodes de multiplication de grands nombres... (voir plus loin).

Une fonction dépendant d'une variable physique t est examinée sur un intervalle de longueur T , soit parce que l'on sait que cette fonction est périodique de période T et est donc entièrement spécifiée par ses valeurs sur un tel intervalle, soit parce que l'on soupçonne qu'elle est de période T , soit enfin parce que l'on juge bon d'étudier *l'extension périodique* de la fonction à partir de ses valeurs sur cet intervalle. On considère donc que $G(t+T) = G(t)$ et le développement

$$G(t) \approx A_0/2 + \sum_{k=1}^{\infty} (A_k \cos(k\omega t) + B_k \sin(k\omega t)),$$

où $\omega = 2\pi/T$ est la **pulsation** du signal, $1/T$ est sa **fréquence**. Ces données s'appliquent à la **fondamentale** $A_1 \cos(\omega t) + B_1 \sin(\omega t)$; chaque **harmonique** $A_k \cos(k\omega t) + B_k \sin(k\omega t)$, $k > 1$, étant de pulsation $k\omega$ et de fréquence k/T .

La justification de ce développement en série sera rappelée plus loin.

Passons à la variable sans dimension $x = 2\pi t/T$ et considérons $F(x) = G(t)$:

$$F(x) \approx A_0/2 + \sum_{k=1}^{\infty} (A_k \cos(kx) + B_k \sin(kx)),$$

passons aux exponentielles complexes $\exp(ikx) = \cos(kx) + i \sin(kx)$:

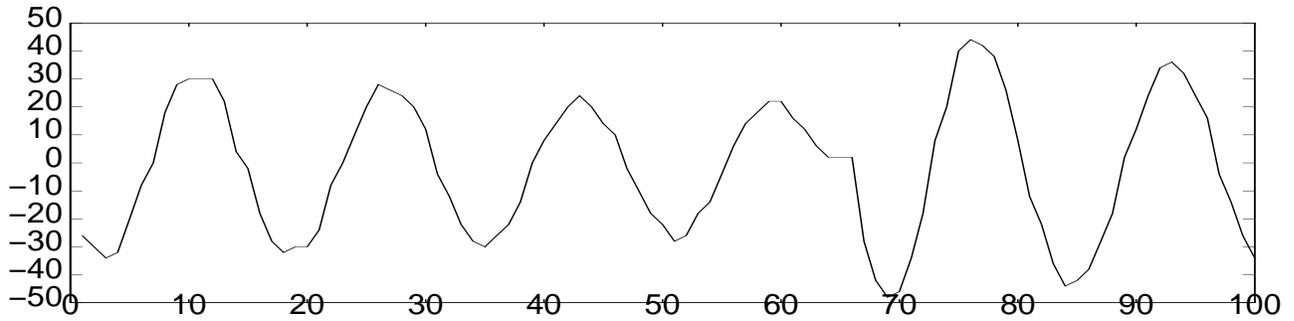
$$F(x) \approx \sum_{k=-\infty}^{+\infty} C_k \exp(ikx),$$

avec $A_k = C_k + C_{-k}$, $B_k = i(C_k - C_{-k})$, $k = 0, 1, \dots$

La figure suivante montre un extrait du fichier "boing.wav", 100 valeurs consécutives prises toutes les 1/11025^{èmes} de seconde d'un signal sonore.

```
%wavfft.m
diary wavfft.dry
nomf = input('nom du fichier wav ', 's')
[y,fs,fm]=wavread(nomf);
fs, fm
[s1,s2]=size(y)
plot(1:s1,y);pause
y2=y(1001:1100)-128;
clear y
plot(1:100,y2);pause;print -dps 'wavfft1';newplot;
```

³¹J. Fourier, *Théorie analytique de la chaleur*, Didot, 1822, = *Analytical Theory of Heat*, (transl. A. Freeman), Dover, 1955.

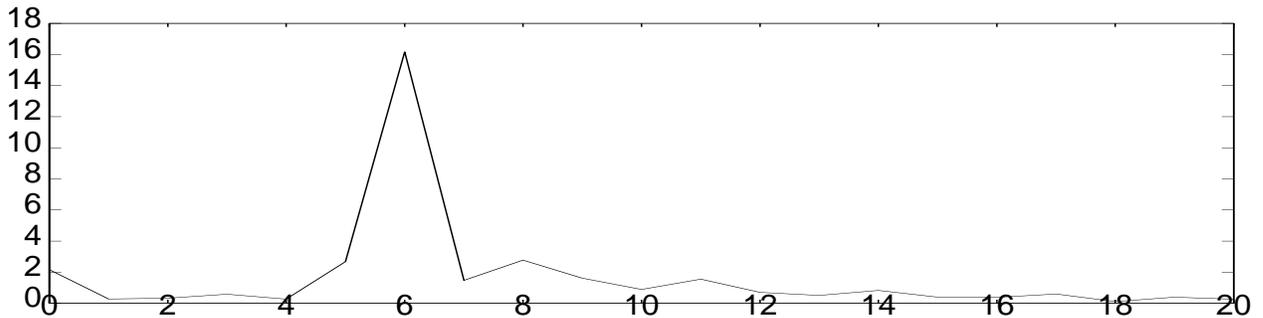


Considérons la fonction constituée de la répétition indéfinie de ces valeurs, c'est donc une fonction périodique de période $T = 100$ pas de temps. On voit aussi que le graphe manifeste une répétition de 6 motifs proches.

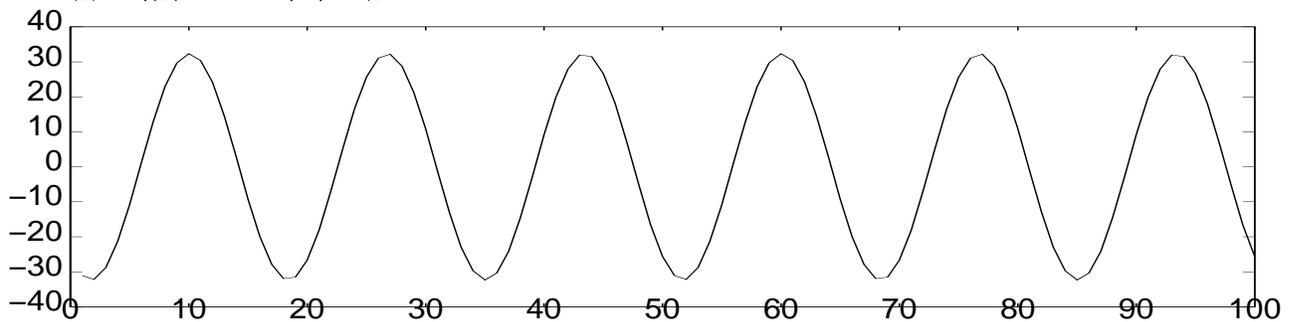
C'est bien ce que montre l'analyse de Fourier de cette fonction:

```
cc=0.01*fft(y2);
cc(1:10)
plot(0:20,abs(cc(1:21)));pause;print -dps 'wavfft2';newplot;
```

$c_0 =$	-2.1600	
$c_1 = \overline{c_{-1}} =$	-0.2343	-0.1470i
$c_2 = \overline{c_{-2}} =$	-0.3038	-0.1246i
$c_3 = \overline{c_{-3}} =$	-0.5831	+ 0.0526i
$c_4 = \overline{c_{-4}} =$	-0.0007	-0.2692i
$c_5 = \overline{c_{-5}} =$	-1.3874	+2.2894i
$c_6 = \overline{c_{-6}} =$	-15.5415	+4.4482i
$c_7 = \overline{c_{-7}} =$	-1.2355	-0.8023i
$c_8 = \overline{c_{-8}} =$	2.2775	-1.5847i
$c_9 = \overline{c_{-9}} =$	0.9349	+1.3177i



On voit que c_6 est nettement plus grand que les autres coefficients. La contribution $c_6 e^{6ix} + c_{-6} e^{-6ix}$

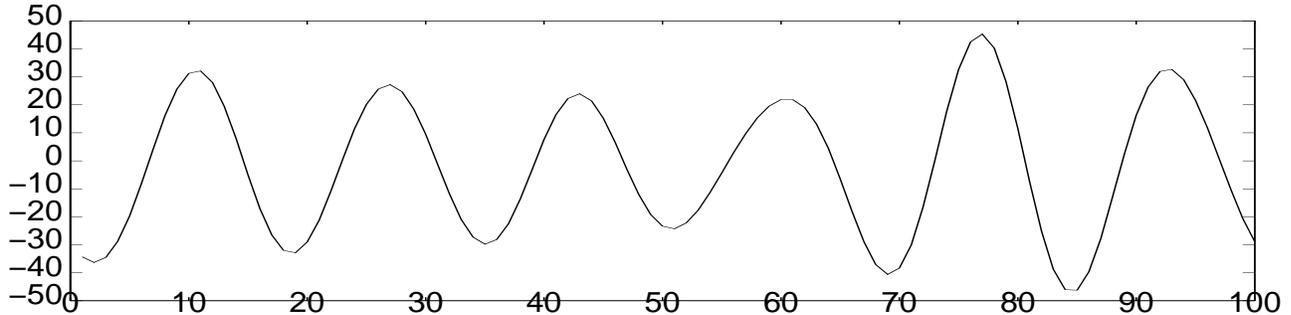


```
cc(1)=cc(1)/2;cc=2*cc;
```

```
dd=0*cc;dd(7)=cc(7);
plot(1:100,100*ifft(dd(1:10),100));print -dps 'wavfft3';newplot
```

La décroissance rapide des coefficients permet de reconstituer la fonction par quelques termes de la série de Fourier:

```
nplot(1:100,100*ifft(cc(1:10),100)):print -dps 'wavfft4'
```



6.1. Comportement des coefficients.

- (1) Si F est intégrable sur $(-\pi, \pi)$: $F \in L^1(-\pi, \pi)$, on a $C_k \rightarrow 0$ quand $k \rightarrow \infty$ (**lemme de Riemann-Lebesgue**).
- (2) Si F est de carré intégrable sur $(-\pi, \pi)$: $F \in L^2(-\pi, \pi)$, les coefficients sont de carré sommable: $\sum_{k=0}^{\infty} |C_k|^2 = (2\pi)^{-1} \|F\|^2 < \infty$ (identité de Parseval).
- (3) Si F est à **variation bornée** sur $[-\pi, \pi]$, c'est-à-dire si $V = \sup_{\theta_j} \sum_j |F(\theta_{j+1}) - F(\theta_j)| < \infty$, $|C_k|$ décroît au moins aussi vite que $1/k$: $|C_k| \leq V/(|k|\pi)$, $|k| > 0$. Par ailleurs, on sait aussi que les sommes de Fourier convergent alors ponctuellement (vers la moyenne des limites à gauche et à droite si on est en un point de discontinuité, théorèmes de **Dirichlet** et **Jordan**).

Examinons maintenant des sous-classes de fonctions continues périodiques:

$$\begin{aligned}
 |k| > 0 : 2\pi C_k &= \int_{-\pi}^{\pi} F(\theta) \exp(-ik\theta) d\theta = \int_{-\pi-\pi/k}^{\pi-\pi/k} F(\theta) \exp(-ik\theta) d\theta = \\
 &= - \int_{-\pi}^{\pi} F(\theta - \pi/k) \exp(-ik\theta) d\theta = \int_{-\pi}^{\pi} \frac{F(\theta) - F(\theta - \pi/k)}{2} \exp(-ik\theta) d\theta
 \end{aligned}$$

d'où

$$|k| > 0 : |C_k| \leq \omega_F(\pi/|k|)/2.$$

- (4) On note Lip_α la classe des fonctions continues telles que $\omega_F(h) \leq Ch^\alpha$, $0 < \alpha \leq 1$ ($\alpha = 1$: classe Lip des fonctions Lipschitziennes)³², on a donc $|C_k| \leq C'/|k|^\alpha$, ce qui n'est pas très brillant (on ne peut avoir $\alpha > 1$ si F n'est pas constante!), mais
- (5) Si $F, F', \dots, F^{(m)}$ continues périodiques (c'est-à-dire continues sur $[-\pi, \pi]$ et $F(-\pi) = F(\pi), \dots, F^{(m)}(-\pi) = F^{(m)}(\pi)$), et si $F^{(m+1)} \in \text{Lip}_\alpha$, $0 < \alpha \leq 1$ (ou $F^{(m+1)}$ à

³²Un exemple intéressant est donné par les séries de Fourier violemment lacunaires $\sum_{m=0}^{\infty} u^m \cos(v^m \theta)$, avec $u > 0$, v entier > 1 , $u < 1$, et surtout $uv > 1$. De telles fonctions sont continues partout et dérivables nulle part (Weierstrass). On calcule qu'elles sont dans Lip_α avec $\alpha = \log(1/u)/\log v$. Si v est impair, les sommes de Fourier sont **aussi** optimales au sens de Tchebycheff (théorème d'équioscillation!).

variation bornée). Alors,

$$|k| > 0 : C_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(\theta) e^{-ik\theta} d\theta = \dots = \frac{1}{2\pi (ik)^{m+1}} \int_{-\pi}^{\pi} F^{(m+1)}(\theta) e^{-ik\theta} d\theta = \frac{1}{(ik)^{m+1}} C_k^{(m+1)},$$

$C_k^{(m+1)}$ étant le coefficient de Fourier de $F^{(m+1)}$, par réduction des termes aux limites apparaissant dans les intégrations par parties. Et comme, par le point précédent, $|C_k^{(m+1)}| \leq C'/|k|^\alpha$,

$$|k| > 0 : |C_k| \leq \frac{C'}{|k|^{m+1+\alpha}},$$

et, selon un raisonnement déjà tenu à propos de séries de polynômes de Tchebycheff:

$$\|F - S_n\|_\infty \leq \sum_{k=n+1}^{\infty} (|C_k| + |C_{-k}|) \leq \frac{2C'}{(m+\alpha)n^{m+\alpha}}.$$

Application. Soit G une fonction de classe \mathcal{C}^2 sur $[a, b]$. On veut approcher G par des combinaisons de fonctions se transformant de façon commode sous l'action d'opérateurs différentiels (méthodes spectrales). C'est le cas des séries de Fourier.

$G((a+b)/2 + (b-a)\theta/(2\pi))$ a cependant un très mauvais développement de Fourier si $G(a) \neq G(b)$. On améliore fortement les choses en procédant comme suit: $t = a + (b-a)\theta/\pi$, $0 \leq \theta \leq \pi$; $G(t) = \xi\theta + \eta + F(\theta)$, où $\xi\theta + \eta$ est l'interpolant linéaire de G en a et b , assurant donc $F(0) = F(\pi) = 0$. On définit alors $F(\theta) = -F(-\theta)$ sur $\theta \in [-\pi, 0]$. Les coefficients de Fourier de F décroissent maintenant comme k^{-3} : $F(\pi) = F(-\pi) = 0$, $F(0_+) = F(0_-) = 0$,

$$F'(0_-) = \lim_{\substack{\theta \rightarrow 0 \\ \theta < 0}} \left(\frac{F(\theta)}{\theta} = \frac{-F(-\theta)}{\theta} = \frac{F(-\theta)}{-\theta} \right) = F'(0_+),$$

$$F'(\pi_-) = \lim_{\substack{\theta \rightarrow 0 \\ \theta > 0}} \left(\frac{F(\pi - \theta)}{-\theta} = \frac{-F(-\pi + \theta)}{-\theta} = \frac{F(-\pi + \theta)}{\theta} \right) = F'((-\pi)_+),$$

donc F et F' sont continues périodiques dans \mathbb{R} : $m = \alpha = 1$.

6.2. Transformée de Fourier discrète. Cf. [GasW]. Soit N pair et considérons le produit scalaire discret

$$(f, g)_N := \sum_{m=0}^{N-1} f(2\pi m/N) \overline{g(2\pi m/N)}.$$

On a alors:

Proposition. Si N est pair, les N fonctions $\exp(ikx)$, $k = -N/2, -N/2 + 1, \dots, N/2 - 2, N/2 - 1$ sont orthogonales relativement au produit scalaire $(\cdot, \cdot)_N$, de même norme \sqrt{N} .

En effet, $(\exp(ik_1x), \exp(ik_2x))_N = \sum_{m=0}^{N-1} \exp(i(k_1 - k_2)2\pi m/N)$, ce qui est une progression géométrique de raison $r = \exp(i(k_1 - k_2)2\pi/N) = 1$ seulement si $k_1 = k_2$, puisque $|k_1 - k_2|$ ne peut valoir d'autre multiple entier de N que 0, et la somme de la progression vaut alors N ; si $k_1 \neq k_2$, la somme vaut $(1 - r^N)/(1 - r) = 0$. \square

La meilleure approximation de f dans l'espace V_N sous-tendu par $\exp(ikx)$, $k = -N/2, -N/2+1, \dots, N/2-2, N/2-1$ est donc

$$f_N(x) = \sum_{k=-N/2}^{N/2-1} c_k^{(N)} \exp(ikx),$$

avec

$$c_k^{(N)} = (f, \exp(ikx))_N / \|\exp(ikx)\|_N^2 = N^{-1} \sum_{m=0}^{N-1} f(2\pi m/N) \exp(-ik2\pi m/N),$$

$$k = -N/2, -N/2+1, \dots, N/2-2, N/2-1.$$

La norme de $f - f_N$ est donc la plus petite norme de $f - p$ que l'on puisse trouver avec $p \in V_N$. Comme cette norme ne fait appel qu'à N valeurs ponctuelles $x = 2\pi m/N$, $m = 0, \dots, N-1$, et que V_N est de dimension N , on s'attend à ce qu'il soit possible de trouver une combinaison p des N fonctions de base de V_N interpolant f aux points $2\pi m/N$, et on aurait donc $\|f - f_N\| \leq \|f - p\| = 0$, ce qui se vérifie:

Proposition. f_N interpole f aux points $x = 2\pi m/N$, $m = 0, \dots, N-1$.

En effet, on vérifie directement que $f_N(2\pi m/N) = \sum_{k=-N/2}^{N/2-1} c_k^{(N)} \exp(ik2\pi m/N)$
 $= N^{-1} \sum_{k=-N/2}^{N/2-1} \sum_{n=0}^{N-1} f(2\pi n/N) \exp(-ik2\pi n/N) \exp(ik2\pi m/N)$
 $= N^{-1} \sum_{n=0}^{N-1} f(2\pi n/N) \sum_{k=-N/2}^{N/2-1} \exp(ik2\pi(m-n)/N)$
 $= N^{-1} \sum_{n=0}^{N-1} f(2\pi n/N) N\delta_{m,n} = f(2\pi m/N)$, ou on applique la remarque de la page 116 et l'identité (79). \square

Cette propriété d'interpolation n'empêche pas les harmoniques de f_N de risquer d'être quelque peu bousculés:

$$\begin{aligned} c_k^{(N)} &= N^{-1} \sum_{m=0}^{N-1} f(2\pi m/N) \exp(-ik2\pi m/N) \\ &= N^{-1} \sum_{m=0}^{N-1} \left[\sum_{\ell=-\infty}^{\infty} c_\ell \exp(i\ell 2\pi m/N) \right] \exp(-ik2\pi m/N) \\ &= N^{-1} \sum_{\ell=-\infty}^{\infty} c_\ell \sum_{m=0}^{N-1} \exp(i(\ell - k)2\pi m/N) \\ &= \dots + c_{k-2N} + c_{k-N} + c_k + c_{k+N} + c_{k+2N} + \dots \end{aligned}$$

Phénomène de pliage fréquentiel, ou "aliasing", qu'on a déjà rencontré avec les développements en séries de polynômes de Tchebycheff. A noter que les harmoniques discrètes sont exactes si le signal initial n'a pas d'harmonique avec $|k| > N/2$ ³³: il faut filtrer un signal avant d'effectuer la discrétisation.

L'analyse d'un signal numérisé consiste donc à calculer les sommes de produits

$$c_k^{(N)} = N^{-1} \sum_{m=0}^{N-1} f(2\pi m/N) \exp(-ik2\pi m/N)$$

³³Ce qu'on peut voir comme une version simplifiée du théorème de Shannon: un signal ne contenant que des fréquences $\leq F$ peut être reconstruit entièrement à partir d'un échantillonnage de pas $\leq 1/(2F)$.

pour N valeurs consécutives de k ;

La synthèse numérique du signal consiste à le reconstituer aux N points $x = 2\pi m/N$, $m = 0, \dots, N - 1$ par calcul de

$$f_N(2\pi m/N) = f(2\pi m/N) = \sum_{k=-N/2}^{N/2-1} c_k^{(N)} \exp(ik2\pi m/N).$$

6.3. Transformée de Fourier rapide (Fast Fourier Transform FFT).

Dans les deux cas, il s'agit d'évaluer N expressions

$$Y_p = \sum_{q=0}^{q=N-1} X_q \zeta^{pq}, \quad (80)$$

où ζ est tantôt $\exp(-2\pi i/N)$ (analyse), tantôt $\exp(2\pi i/N)$ (synthèse). Il nous importe seulement que $\zeta^N = 1$. Même si les ζ^{pq} sont précalculés, il semble inévitable de devoir effectuer N additions et N multiplications par Y_p , soit $2N^2$ opérations en tout.

Le caractère relatif de cette obligation apparente apparaît mieux quand on examine des problèmes multidimensionnels: ainsi, un problème à deux dimensions demande de calculer

$$Y_{p_1, p_2} = \sum_{q_1=0}^{q_1=N_1-1} \sum_{q_2=0}^{q_2=N_2-1} X_{q_1, q_2} \zeta_1^{p_1 q_1} \zeta_2^{p_2 q_2},$$

pour $N_1 N_2$ couples (p_1, p_2) , d'où apparemment $2(N_1 N_2)^2$ opérations. Mais on voit que si on écrit

$$Y_{p_1, p_2} = \sum_{q_1=0}^{q_1=N_1-1} \zeta_1^{p_1 q_1} \left(\sum_{q_2=0}^{q_2=N_2-1} X_{q_1, q_2} \zeta_2^{p_2 q_2} \right),$$

on évalue d'abord les $N_1 N_2$ expressions intermédiaires $Z_{p_2, q_1} = \sum_{q_2=0}^{q_2=N_2-1} X_{q_1, q_2} \zeta_2^{p_2 q_2}$ en $2(N_1 N_2) N_2$ opérations, et on obtient le résultat final $Y_{p_1, p_2} = \sum_{q_1=0}^{q_1=N_1-1} \zeta_1^{p_1 q_1} Z_{p_2, q_1}$ en $2(N_1 N_2) N_1$ opérations supplémentaires, soit $2(N_1 N_2)(N_1 + N_2)$ opérations en tout!

Avec un problème tridimensionnel, on aurait $2(N_1 N_2 N_3)(N_1 + N_2 + N_3)$ opérations.

L'algorithme de transformée discrète de Fourier rapide (fast Fourier transform **FFT**) de Cooley et Tukey³⁴ introduit systématiquement ce mécanisme:

si N n'est pas un nombre premier, soit $N = N_1 N_2$, effectuons la division de $p \in \{0, 1, \dots, N - 1\}$ par N_2 , soit p_2 le reste $\in \{0, 1, \dots, N_2 - 1\}$ et p_1 le quotient $\in \{0, 1, \dots, N_1 - 1\}$ puisque $p < N$. De même, on effectue la division de q par N_1 : $q = q_2 N_1 + q_1$, $0 \leq q_1 < N_1$, $0 \leq q_2 < N_2$, et (80) devient

$$\begin{aligned} Y_p &= Y_{p_1 N_2 + p_2} = \sum_{\substack{0 \leq q_1 < N_1 \\ 0 \leq q_2 < N_2}} X_{q_2 N_1 + q_1} \zeta^{(p_1 N_2 + p_2)(q_2 N_1 + q_1)} \\ &= \sum_{\substack{0 \leq q_1 < N_1 \\ 0 \leq q_2 < N_2}} X_{q_2 N_1 + q_1} \zeta^{p_1 q_2 N_2 N_1 + p_1 q_1 N_2 + p_2 q_2 N_1 + p_2 q_1} \\ &= \sum_{\substack{0 \leq q_1 < N_1 \\ 0 \leq q_2 < N_2}} X_{q_2 N_1 + q_1} \zeta^{p_1 q_1 N_2 + p_2 q_2 N_1 + p_2 q_1} \quad \text{puisque } \zeta^{N_1 N_2} = \zeta^N = 1, \\ &= \sum_{q_1=0}^{N_1-1} \left[\sum_{q_2=0}^{N_2-1} X_{q_2 N_2 + q_1} (\zeta^{N_1})^{p_2 q_2} \right] \zeta^{p_2 q_1} (\zeta^{N_2})^{p_1 q_1}. \end{aligned}$$

³⁴J.W. Cooley, J.W. Tukey, An algorithm for the machine calculation of complex Fourier series, *Math. Comput.* **19** (1965) 297-301. Cet article, arrivé à point nommé, est crédité de 2155 citations fin 1993, un record. Après coup, on a retrouvé des précurseurs significatifs dans des ouvrages de Lanczos, Runge, et même Gauss (cf. pp. 8 et 9 de *Current Contents PC&ES* **33** (20-27 Dec. 1993) #51-52).

On retrouve la forme bidimensionnelle. Ce n'est pas tout! Les expressions intérieures Z_{p_2, q_1} (entres crochets) sont encore de la forme (80), avec ζ^{N_1} au lieu de ζ , on doit donc pouvoir les évaluer à leur tour par FFT si N_2 n'est pas premier (N est remplacé par N_2 : on a bien $(\zeta^{N_1})^{N_2} = 1$), etc. Et le résultat final est lui-même une FFT de taille N_1 réalisée avec ζ^{N_2} .

On a donc la description récursive suivante:

Algorithme FFT:

- Si N est premier, effectuer (80) en $2N^2$ opérations.
- Si $N = N_1 N_2$,
 Pour $q_1 = 0, \dots, N_1 - 1$:
 – Effectuer l'algorithme FFT avec les N_2 données $X_{q_2 N_1 + q_1}$, $q_2 = 0, \dots, N_2 - 1$. On recueille ainsi un vecteur d'éléments Z_{p_2, q_1} , $p_2 = 0, \dots, N_2 - 1$.
 Fin de la boucle q_1 .
- On multiplie chaque Z_{p_2, q_1} par le "twiddle factor" $\zeta^{p_2 q_1}$.
- Pour $p_2 = 0, \dots, N_2 - 1$:
 – Effectuer l'algorithme FFT avec les N_1 données $Z_{p_2, q_1} \zeta^{p_2 q_1}$, $q_1 = 0, \dots, N_1 - 1$. On recueille ainsi $Y_{p_1 N_2 + p_2}$, $p_1 = 0, \dots, N_1 - 1$.
 Fin de la boucle p_2 .

Apprécions maintenant le nombre \mathcal{N}_N d'opérations réellement effectuées: si N n'est pas premier, $\mathcal{N}_N = N_1 \mathcal{N}_{N_2} + N + N_2 \mathcal{N}_{N_1}$, ou encore:

$$\frac{\mathcal{N}_N}{N} = \frac{\mathcal{N}_{N_1}}{N_1} + \frac{\mathcal{N}_{N_2}}{N_2} + 1.$$

Si ρ_1, \dots, ρ_ν sont les facteurs premiers (pas nécessairement distincts) de N , on a donc

$$\mathcal{N}_N = N \left(\nu + \sum_{\mu=1}^{\nu} \frac{\mathcal{N}_{\rho_\mu}}{\rho_\mu} \right) = N \left(\nu + 2 \sum_{\mu=1}^{\nu} \rho_\mu \right).$$

En particulier, si N est une puissance de 2, $N = 2^\nu$,

$$\mathcal{N}_N = 5N\nu = 5N \log_2 N,$$

en fait, encore moins, puisque certains "twiddle factors" valent 1.

Exercice. Reprendre le raisonnement avec N pair, $N_1 = 2$, $N_2 = N/2$:

$$\begin{array}{rcll} Y_0 & = & X_0 + X_2 + \dots + X_{N-2} + X_1 + X_3 + \dots + X_{N-1} \\ Y_1 & = & X_0 + X_2 \zeta^2 + \dots + X_{N-2} \zeta^{N-2} + \zeta (X_1 + X_3 \zeta^2 + \dots + X_{N-1} \zeta^{N-2}) \\ Y_2 & = & X_0 + X_2 \zeta^4 + \dots + X_{N-2} \zeta^{2N-4} + \zeta^2 (X_1 + X_3 \zeta^4 + \dots + X_{N-1} \zeta^{2N-4}) \\ \dots & & \dots & \dots & \dots & \dots \\ Y_{N/2} & = & X_0 + X_2 + \dots + X_{N-2} - (X_1 + X_3 + \dots + X_{N-1}) \\ Y_{N/2+1} & = & X_0 + X_2 \zeta^2 + \dots + X_{N-2} \zeta^{N-2} - \zeta (X_1 + X_3 \zeta^2 + \dots + X_{N-1} \zeta^{N-2}) \\ Y_{N/2+2} & = & X_0 + X_2 \zeta^4 + \dots + X_{N-2} \zeta^{2N-4} - \zeta^2 (X_1 + X_3 \zeta^4 + \dots + X_{N-1} \zeta^{2N-4}) \\ \dots & & \dots & \dots & \dots & \dots \end{array}$$

où on utilise $\zeta^{N/2} = -1$.

Convolutions et produits de grands nombres.

La **convolution** de deux suites $\{x_0, x_1, \dots, x_{N-1}\}$ et $\{y_0, y_1, \dots, y_{N-1}\}$ est la suite $\{z_k = \sum_{0 \leq \ell < N, 0 \leq m < N, \ell+m=k} x_\ell y_m\}$, $k = 0, 1, \dots, 2N - 2$. Le calcul de $z_0, z_1, \dots, z_{2N-2}$ prend apparemment $2N^2$ opérations. *Aha!*

Les z_k sont effectivement les coefficients de

$$f(u)g(u) = \left(\sum_{\ell=0}^{N-1} x_\ell u^\ell \right) \left(\sum_{m=0}^{N-1} y_m u^m \right) = \sum_{k=0}^{2N-2} z_k u^k.$$

Il suffit donc d'évaluer f et g en $u = 1, \zeta, \zeta^2, \dots, \zeta^{2N-1}$ (**synthèse**), avec $\zeta = \exp(\pi i/N)$ (on travaille avec $2N$ au lieu de N), et de récupérer les z_k à partir de ces $2N$ valeurs (**analyse**), donc, tout cela avec trois FFT! ³⁵

³⁵Cf. P. Henrici, *Applied & Computational Complex Analysis III*, Wiley, 1986, et Fast Fourier methods in computational complex analysis, *SIAM Rev.* **21** (1979) 481-527.

Produit de grands nombres: si F et G sont donnés par leurs développements en base b $F = \sum_{\ell=0}^{N-1} x_{\ell} b^{\ell}$, $G = \sum_{m=0}^{N-1} y_m b^m$, on obtient facilement le développement en base b de FG ³⁶

Karatsuba: $ab = [(a+b)^2 - (a-b)^2]/4$, il suffit de savoir évaluer des carrés. Soit X un nombre de p chiffres en base b , on écrit $X = Yb^{p/2} + Z$, et X^2 fait appel à Y^2 , Z^2 , YZ évalué par $(Y \pm Z)^2$. MAIS $(Y - Z)^2 = 2Y^2 + 2Z^2 - (Y + Z)^2$! Trois carrés de nombres de $p/2$ chiffres suffisent. Coût $C(p) = 3C(p/2) + \text{const.} \cdot p$. Cela donne $p(3/2)^{\log_2 p} = p^{1+\log_2(3/2)} = p^{1.585}$.

Voir aussi

NFFT, Nonequispaced Discrete Fourier Transform

For those who would prefer to experiment with such tools in Matlab, we have independently developed a NUFFT toolbox that uses interpolators that have been min-max optimized to minimize the worst-case interpolation error. The toolbox is one part of a large collection of m-files developed for image reconstruction problems, and is located here:

<http://www.eecs.umich.edu/~fessler/code/index.html>

Sincerely, Jeff Fessler

...

The NFFT has a lot of applications, see e.g.

<http://www.math.uni-luebeck.de/potts/nfft/links.sql>

such as: - summation at nonequispaced knots, evaluation of radial functions - spherical Fourier algorithms - Fourier reconstruction algorithms for (medical) imaging CT/MRI - a new method for particle simulations

From: Steven G. Johnson <stevenj@fftw.org> Date: Fri, 14 May 2004 21:10:36 -0400 (EDT) Subject: Harminv 1.0, Extracting Frequencies Better Than the FFT

ANNOUNCE: Harminv 1.0 – extracting frequencies better than the FFT

I'm pleased to announce the availability of Harminv 1.0, a free program and C library for harmonic inversion: decomposing a time-series into a sum of sinusoids, including exponentially decaying sinusoids.

<http://ab-initio.mit.edu/harminv/>

Harminv is an implementation of the "filter diagonalization method" (FDM) of Mandelshtam & Taylor (see URL for references), which maps the harmonic inversion problem onto a small eigen-problem (size proportional to the number of sinusoids).

This method has been widely employed in physics since its inception in 1997, and is often able to obtain much more robust and accurate solutions for the frequencies, etcetera, than e.g. direct least-squared fitting of the data or its FFT spectrum (which are typically very ill-conditioned approaches).

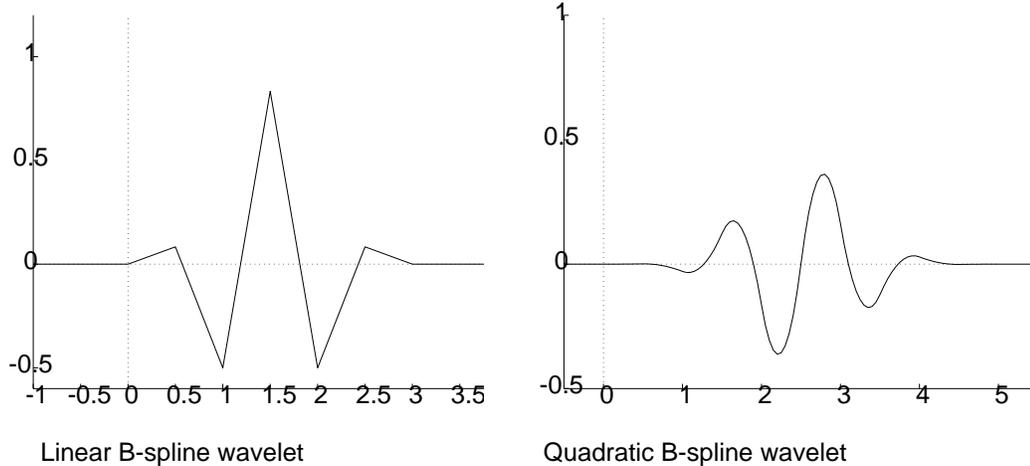
I've been happily using this code for a few years now in my own research, and I finally got around to releasing it. I hope that others find it useful.

Cordially, Steven G. Johnson

6.4. Analyse en ondelettes.

Soit ψ une fonction de support compact. On décide de représenter des fonctions définies sur \mathbb{R} en une combinaison de fonctions $\psi_{a,b}(x) = |a|^{-1/2} \psi\left(\frac{x}{a} - b\right)$ (**ondelettes**) pour un ensemble dénombrable de paramètres a et b . ψ est une fonction qui oscille plusieurs fois sur un support borné:

³⁶Technique effectivement utilisée dans D.H. Bailey, Algorithm 719. Multiprecision translation and execution of FORTRAN programs, *ACM Trans. Math. Soft.* **19** (1993) 288-319.



Exemples d'ondelettes (fonctions ψ), extrait de M. Ueda, S. Lodha, Wavelets: An Elementary Introduction and Examples, <ftp://ftp.cse.ucsc.edu/pub/tr/ucsc-crl-94-47.ps.Z>

Le coefficient de $\psi_{a,b}$ d'une fonction nous renseigne sur l'amplitude de cette fonction dans un voisinage de ab , information qu'on ne retrouve pas avec des séries ou transformées de Fourier; cependant, pour a petit, le coefficient de $\psi_{a,b}$ fournit une information fréquentielle, l'ondelette se comportant alors comme un train d'ondes de haute fréquence. On espère ainsi disposer d'un outil très souple d'analyse de signal.

Pour a donné, on note V_a l'espace vectoriel sous-tendu par des fonctions φ_{a,kb_0} , $k \in \mathbb{Z}$. V_a sera par exemple l'ensemble des fonctions continues linéaires par morceaux dans les intervalles $[kab_0, (k+1)ab_0]$, et φ_{a,kb_0} sera alors $1 - |x - kab_0|/(ab_0)$ dans $[(k-1)ab_0, (k+1)ab_0]$, 0 ailleurs (ce ne sont pas encore les fonctions ψ). On considère une suite $\{V_{a_j}\}$ de tels espaces permettant donc de représenter de plus en plus finement une fonction à mesure que a_j est petit (*analyse multi-résolution*). Les approximations successives f_j et f_{j+1} sont les projections orthogonales de f dans V_{a_j} et $V_{a_{j+1}}$. Les **espaces d'ondelettes** W_j contiennent les différences (**détails**) $f_{j+1} - f_j$:

$$\begin{aligned}
 V_{a_{j+1}} &= V_{a_j} \oplus W_j. \\
 f_j(x) &= \sum_{k=-\infty}^{\infty} \lambda_{j,k} \varphi_{a_j, kb_0}(x), \\
 f_{j+1}(x) &= \sum_{k=-\infty}^{\infty} \lambda_{j+1,k} \varphi_{a_{j+1}, kb_0}(x), \\
 &= f_j(x) + \sum_{k=-\infty}^{\infty} \gamma_{j,k} \psi_{a_j, kb_0}(x).
 \end{aligned}$$

Et donc, formellement,

$$f(x) = \sum_{j=-\infty}^{\infty} \left[\sum_{k=-\infty}^{\infty} \gamma_{j,k} \psi_{a_j, kb_0}(x) \right].$$

On s'arrange généralement pour que les W_j soient orthogonaux entre eux: $\psi_{a_p, b} \perp \psi_{a_q, c}$ si $p \neq q$, parfois pour que toutes les $\psi_{a,b}$ soient orthogonales entre elles (plus rare).

Exemple de détermination de ψ à partir de φ : soit $a_j = a_0/2^j$, $V_{a_0} =$ l'espace des fonctions linéaires par morceaux sur les intervalles $[ka_0b_0, (k+1)a_0b_0]$, $\varphi(x) = 1 - |x|/b_0$ sur $[-b_0, b_0]$, $\varphi(x) = 0$ quand $|x| > b_0$. $\psi(x/a_0)$ est une fonction de $V_{a_1} = V_{a_0/2}$ orthogonale à toutes les fonctions $\varphi_{a_0, kb_0}(x) = \varphi(x/a_0 - kb_0)$. On trouve

$$\psi((k-1)b_0) + 6\psi((k-1/2)b_0) + 10\psi(kb_0) + 6\psi((k+1/2)b_0) + \psi((k+1)b_0) = 0, \quad k \in \mathbb{Z}.$$

Il y a plusieurs solutions de support borné, par exemple (figure),

$$\psi(b_0/2) = 1, \psi(b_0) = -6, \psi(3b_0/2) = 10, \psi(2b_0) = -6, \psi(5b_0/2) = 1, \psi(x) = 0 \text{ si } x \leq 0 \text{ ou } x \geq 3b_0.$$

Nombreuses applications en traitement d'informations visuelles et sonores, <http://www.larecherche.fr/arch/02/01> etc.

Ref: [GasW], nombreux livres et articles de I. Daubechies, Y. Meyer, etc., nombreuses contributions sur le web, par exemple <http://www.cs.kuleuven.ac.be/~wavelets/>

P. Bechler, Lebesgue constant for the Strömberg wavelet, *J. Approx. Theory* **122** (2003) 13-23.

From: Brigitte Forster <brigitte.forster@epfl.ch> Date: Mon, 27 Jan 2003 09:04:40 +0100 Subject: Revamped Wavelet Digest

Dear colleagues,

We are very happy to announce you that the revamped wavelet digest took off.

<http://www.wavelet.org/>

The Wavelet Digest was founded by Wim Sweldens. Its first issue was released on July 24, 1992. Since then, the impact of the digest on the community kept increasing and the number of subscribers grew up to almost 20000 by the end of 2001. During all those years, Wim has edited and assembled the digest. Now, the Wavelet Digest takes off again, restyled and hosted by the Swiss Federal Institute of Technology Lausanne (EPFL).

So if you have any information you want to send out to the wavelet community, please submit it to the digest. The digest will be send out depending on the number of submissions. We do hope to have an issue about every two weeks.

To subscribe the wavelet digest please follow the link <http://www.wavelet.org/index.php?subscribe=1>

With best regards, The Wavelet Digest Team

From: Laurent Demanet <demanet@acm.caltech.edu> Date: Thu, 14 Jul 2005 Subject: The Curvelab Toolbox at Curvelet.org

Curvelet is a new toolbox implementing the Curvelet transform, both in Matlab and C++, and can be downloaded from <http://www.curvelet.org>

Curvelets are multiscale oriented basis elements that were introduced by Candes and Donoho to address the problem of representation of edges in geometrical images. Since then, they have found applications at least in seismic imaging, astronomy, and the numerical analysis of wave equations.

The software Curvelet contains two distinct implementations of the 2D Curvelet transform, namely via Wrapping and via USFFT. Both architectures are introduced and explained in the report 'Fast Discrete Curvelet Transforms', available online. The 3D transform is also present in Curvelet, in three different versions: in-core, out-of-core and fully MPI-based parallel. The 3D code is covered in a separate online report. Several demo files illustrate the transforms at work on image processing tasks such as denoising and partial reconstructions. Additional routines are provided to help the user understand the geometry of the transform (location, orientation and scale for each coefficient.) A user's guide explains how to set up the toolbox step-by-step on your computer.

The webpage <http://www.curvelet.org> is also meant to serve as a repository of papers related to curvelets as implemented in Curvelet. It contains a list of links to applications of curvelets, and researchers in the field. There is also a mailing list: sign up to stay up-to-date on future software releases and other curvelet-related information.

The Curvelet team: E. Candes, L. Demanet, D. Donoho and L. Ying.

7. Convergence, espace de Hilbert

7.1. Suites totales et maximales.

Soit $\{\varphi_0, \varphi_1, \dots\}$ une suite orthonormale de l'espace préhilbertien X sur \mathbb{C} , et V_n le sous-espace de base $\{\varphi_0, \dots, \varphi_n\}$, $n = 0, 1, \dots$

Comme $V_n \subset V_{n+1}$, la norme d'erreur de meilleure approximation de $f \in X$ dans V_n décroît quand n augmente:

$$\hat{p}_n = \sum_{k=0}^n c_k \varphi_k, \quad \|f - \hat{p}_n\|^2 = \|f\|^2 - \sum_{k=0}^n |c_k|^2,$$

où $c_k = (f, \varphi_k)$, $k = 0, 1, \dots$

La suite des coefficients c_k de tout $f \in X$ est donc toujours de carré sommable, et

$$\forall f \in X, \quad \sum_{k=0}^{\infty} |c_k|^2 \leq \|f\|^2,$$

(*inégalité de Bessel*).

Définition. Une suite $\{\varphi_0, \varphi_1, \dots\}$ d'un espace normé X est **totale** dans X si les combinaisons linéaires finies des éléments de la suite forment un ensemble dense dans X , c'est-à-dire si, $\forall \varepsilon > 0$, $\exists p = c_0 \varphi_0 + \dots + c_n \varphi_n$: $\|f - p\| \leq \varepsilon$.

On va donc devoir se poser la question de savoir si $\{\varphi_0, \varphi_1, \dots\}$ est totale dans notre espace X donné.

Existe-t-il $f \in X$ qu'on ne pourrait **pas** approcher d'aussi près que l'on veut par des combinaisons des φ_k ? Les meilleures combinaisons possibles sont les $\hat{p}_n = \sum_{k=0}^n c_k \varphi_k$. Quand n augmente, les $f - \hat{p}_n$, qui gardent des normes ne tendant pas vers zero, sont orthogonaux à de plus en plus de φ_k , $k = 0, 1, \dots$. A la limite (tout est dans ce mot), $f - \hat{p}_\infty$ serait orthogonal à **tous** les φ_k .

Pour exploiter cette idée de possibilité, ou d'impossibilité, d'existence d'un élément non nul de X orthogonal à tous les φ_k , il faut donner un sens à $\sum_{k=0}^{\infty} c_k \varphi_k$.

Exemple. Prenons $X =$ espace des fonctions mesurables bornées sur $[-\pi, \pi]$ (c'est-à-dire $\forall f \in X, \exists M < \infty : |f(x)| \leq M, -\pi \leq x \leq \pi$), muni du produit scalaire habituel³⁷ $(f, g) = \int_{-\pi}^{\pi} f(x)g(x)dx$. La suite $\varphi_k(x) = (2\pi)^{-1/2} \exp(ikx)$, $k = 0, 1, \dots$ est une suite orthonormale de X . Avec $f(x) = \text{sign } x$, on trouve $c_k = 0$ si k est pair, $c_k = -4i(2\pi)^{-1/2}/k$ si k est impair, mais $\sum_{k=1}^n c_k \varphi_k^*(x) \rightarrow \text{sign } x - 2i\pi^{-1} \ln |\cot(x/2)|$ quand $n \rightarrow \infty$, ce qui n'est pas borné, donc pas dans X .

On montrera plus loin que, si la suite $\varphi_k(x) = (2\pi)^{-1/2} \exp(ikx)$, $k = 0, 1, \dots$ n'est effectivement pas totale, la suite $\varphi_{2k}(x) = (2\pi)^{-1/2} \exp(ikx)$, $k = 0, 1, \dots$, $\varphi_{2k+1}(x) = (2\pi)^{-1/2} \exp(-ikx)$, $k = 0, 1, \dots$, l'est.

Proposition. Pour toute suite $\{c_k\}_0^\infty$ de carré sommable $\sum_0^\infty |c_k|^2 < \infty$, la suite $\{\hat{p}_n = \sum_{k=0}^n c_k \varphi_k\}_0^\infty$ est une suite de Cauchy dans X .

³⁷Une fonction mesurable de valeur absolue bornée est intégrable; le produit de deux fonctions de ce type possède encore les mêmes propriétés, et est donc intégrable

En effet, vérifions que, $\forall \varepsilon > 0$, on peut trouver n tel que $\|\hat{p}_{n_2} - \hat{p}_{n_1}\| \leq \varepsilon$ dès que $n_2 \geq n_1 \geq n$: il suffit de remarquer que $\|\hat{p}_{n_2} - \hat{p}_{n_1}\|^2 = (\hat{p}_{n_2} - \hat{p}_{n_1}, \hat{p}_{n_2} - \hat{p}_{n_1}) = (\sum_{n_1+1}^{n_2} c_n \varphi_n, \sum_{n_1+1}^{n_2} c_n \varphi_n) = |c_{n_1+1}|^2 + \dots + |c_{n_2}|^2$. \square

Définition. Un **espace de Hilbert** est un espace préhilbertien³⁸ **complet** (toute suite de Cauchy dans X a donc une limite dans X). Un espace de Hilbert **séparable** est un espace de Hilbert admettant une suite totale (suite = famille dénombrable).

Définition. Une suite orthonormale d'un espace préhilbertien X est **maximale**³⁹ si on ne peut trouver de suite orthonormale plus grande dans X .

Nous avons maintenant tous les éléments pour apprécier le

7.2. Théorème.

Soit $\{\varphi_0, \varphi_1, \dots\}$ une suite orthonormale de l'espace préhilbertien X . Considérons les 7 propositions:

- (1) La suite $\{\varphi_0, \varphi_1, \dots\}$ est totale dans X .
- (2) Le développement en série des φ_k de tout $f \in X$ converge en norme vers f :

$$\lim_{n \rightarrow \infty} \|f - \hat{p}_n\| = \|f - \sum_{k=0}^n (f, \varphi_k) \varphi_k\| = 0,$$

- (3) Pour tout $f \in X$, on a

$$\|f\|^2 = \sum_{k=0}^{\infty} |c_k|^2 = \sum_{k=0}^{\infty} |(f, \varphi_k)|^2,$$

(**relation de Parseval**),

- (4) Pour tout $f, g \in X$, on a

$$(f, g) = \sum_{k=0}^{\infty} c_k \overline{d_k} = \sum_{k=0}^{\infty} (f, \varphi_k) (\varphi_k, g),$$

(**relation de Parseval étendue, ou Riesz-Fischer**),

- (5) La suite $\{\varphi_0, \varphi_1, \dots\}$ est maximale dans X .
- (6) $f \in X$, $(f, \varphi_k) = 0$, $k = 0, 1, \dots \Rightarrow f = 0$,
- (7) Tout élément f de X est entièrement déterminé par ses coefficients $c_k = (f, \varphi_k)$, $k = 0, 1, \dots$: $(f, \varphi_k) = (g, \varphi_k)$, $k = 0, 1, \dots \Rightarrow f = g$.

Alors,

$$(1) \Leftrightarrow (2) \Leftrightarrow (3) \Leftrightarrow (4) \Rightarrow (5) \Leftrightarrow (6) \Leftrightarrow (7).$$

De plus, **si X est un espace de Hilbert**, les propositions (1) à (7) sont équivalentes.

³⁸Rappelons qu'un espace préhilbertien est considéré ici comme normé, donc séparé, cf. § 1.2, p. 76.

³⁹Cf. [Dav] pp.191–194, aussi p.257. Attention! Davis appelle “closed” ce qui est ici “totale”, et “complete” ce qui est ici “maximale”. Pour ne rien arranger, R. Courant, D. Hilbert, *Methods of Mathematical Physics I*, Interscience 1953, adopte la terminologie inverse.

Pour la nomenclature en français, cf. par exemple, J.-P. Bertrandias, *Mathématique pour l'informatique. 1-Analyse fonctionnelle*, A. Colin, 1970, chap. 3, G. Choquet, *Cours d'analyse II, topologie*, Masson, 1969, chap. VII-IV, J. Dieudonné, *Fondements de l'analyse moderne*, Gauthier-Villars, 1965, chap. 6.

En effet, on a déjà établi l'équivalence entre (1), (2) et (3), puisque les \hat{p}_n sont les meilleures combinaisons possibles, et $\|f - \hat{p}_n\|^2 = \|f\|^2 - \sum_{k=0}^n |c_k|^2$.

Montrons que (2) \Rightarrow (4): comme $f - \hat{p}_n$ et $g - \hat{q}_n$ sont orthogonaux à V_n ,

$$(f - \hat{p}_n, g - \hat{q}_n) = (f - \hat{p}_n, g) = (f, g) - (\hat{p}_n, g) = (f, g) - (\hat{p}_n, g - \hat{q}_n + \hat{q}_n) = \\ (f, g) - (\hat{p}_n, \hat{q}_n) = (f, g) - \sum_{k=0}^n c_k \overline{d_k},$$

Appliquons l'inégalité de Cauchy-Schwarz:

$$|(f, g) - \sum_{k=0}^n c_k \overline{d_k}| = |(f - \hat{p}_n, g - \hat{q}_n)| \leq \|f - \hat{p}_n\| \cdot \|g - \hat{q}_n\|$$

qui tend bien vers 0 quand $n \rightarrow \infty$, par (2). (4) \Rightarrow (3): il suffit de prendre $g = f$.

(1) \Rightarrow (5): supposons que l'on puisse construire $\psi \in X$, $\|\psi\| = 1$, tel que la suite $\{\psi, \varphi_0, \varphi_1, \dots\}$ soit encore orthonormale dans X . Mais $\{\varphi_0, \varphi_1, \dots\}$ étant totale, appliquons Parseval avec $f = \psi$: $\|\psi\|^2 = \sum_{k=0}^{\infty} |c_k|^2 = \sum_{k=0}^{\infty} |(\psi, \varphi_k)|^2 = 0$, contradiction.

(5) \Leftrightarrow (6) est immédiat.

(6) \Rightarrow (7): on utilise (6) avec $f - g$ au lieu de f ; (7) \Rightarrow (6): on prend $g = 0$.

Soit X Hilbert, montrons que l'une des propositions (5) – (7) implique l'une des propositions (1) – (4), choisissons (7) \Rightarrow (2): les approximations \hat{p}_n de f forment une suite de Cauchy et tendent donc vers un certain $p \in X$. Il faut montrer que $p = f$. Or, pour tout k fixé, $(p, \varphi_k) = \lim_{n \rightarrow \infty} (\hat{p}_n, \varphi_k) = (f, \varphi_k)$ à partir de $n = k$, donc p et f ont tous leurs coefficients identiques, donc $p = f$. \square

Contre-exemple. La fonction $f(x) = \sin(x^\alpha \operatorname{tg}(\alpha\pi))$ est orthogonale à tous les polynômes selon le produit scalaire $(f, g) = \int_0^\infty f(x)g(x) \exp(-x^\alpha) dx$ si $0 < \alpha < 1/2$ (Stieltjes, Hamburger, cité dans O. Perron, *Die Lehre von den Kettenbrüchen*, 1929, § 67). On applique une identité de la fonction Gamma (déformation du parcours d'intégration) $\int_0^\infty t^{\nu-1} \exp(-te^{i\beta}) dt = \exp(i\nu\beta)\Gamma(\nu)$ si $\nu > 0$ et $-\pi/2 < \beta < \pi/2$. La partie imaginaire est nulle si $\nu\beta$ est un multiple entier > 0 de π : soit $\beta = \alpha\pi$ et $\nu = n/\alpha$, $n = 1, 2, \dots$, et on pose $t = x^\alpha / \cos(\alpha\pi)$. On obtient bien $\int_0^\infty f(x)x^{n-1} \exp(-x^\alpha) dx = 0$ pour $n - 1 = 0, 1, \dots$. La suite des polynômes est donc non maximale, donc non totale, dans tout préhilbertien pour le produit scalaire indiqué, et contenant f .

7.3. Exemples de suites totales dans $\mathcal{C}[a, b]$ et $L^2([a, b], \mu)$.

$L^2([a, b], \mu)$ est l'espace des fonctions de carré μ -intégrables sur $[a, b]$, muni du produit scalaire $(f, g) = \int_a^b f(x)\overline{g(x)} d\mu(x)$, f et g étant considérées comme équivalentes si $\|f - g\| = 0$. Cet espace de fonctions (en fait, de classes d'équivalence de fonctions) est un espace de Hilbert. Lorsque $[a, b]$ est borné, l'espace des fonctions continues $\mathcal{C}[a, b]$ est dense dans $L^2([a, b], \mu)$ (au sens de la norme de L^2), ce qui permet d'établir le caractère total de suites de L^2 en passant par \mathcal{C} , ce qui est d'ailleurs très intéressant en soi.

7.4. Théorème d'approximation de Weierstrass.

Si $[a, b]$ est borné, les polynômes forment un ensemble dense dans $\mathcal{C}[a, b]$, muni de la norme du maximum.

La démonstration utilisera une construction très utile et très ingénieuse à de nombreux égards:

Polynômes de Bernstein.

Soit f une fonction définie sur $[0, 1]$. On considère l'opérateur \mathcal{B}_n :

$$(\mathcal{B}_n f)(t) := \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} t^k (1-t)^{n-k},$$

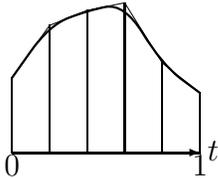
où $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ est un coefficient binomial: $(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$. $\binom{n}{k}$ est

le nombre de combinaisons sans répétition de n objets pris k à k : $\binom{n}{0} = 1$, $\binom{n}{1} = n$,

$$\binom{n}{2} = n(n-1)/2, \dots, \binom{n}{n} = 1, \binom{n}{k} = \binom{n}{n-k}.$$

On constate rapidement $(\mathcal{B}_n f)(0) = f(0)$, $(\mathcal{B}_n f)(1) = f(1)$, $f(t) \equiv 1 \Rightarrow (\mathcal{B}_n f)(t) \equiv 1$, \mathcal{B}_n est un opérateur **linéaire positif**, on entend par ce dernier terme que $f(t) \geq 0$ sur $[0, 1] \Rightarrow (\mathcal{B}_n f)(t) \geq 0$ sur $[0, 1]$.

Remarquons que la linéarité et la positivité impliquent $|f(t)| \leq g(t)$ sur $[0, 1] \Rightarrow |(\mathcal{B}_n f)(t)| \leq (\mathcal{B}_n g)(t)$ sur $[0, 1]$.



Pour aller plus loin, constatons que $w_k(t) := \binom{n}{k} t^k (1-t)^{n-k}$ est la probabilité d'obtenir k fois un évènement de probabilité t sur n tirages indépendants: **loi binomiale**.

Des propriétés bien connues de cette loi sont:

- (1) Prob. totale $\sum_{k=0}^n w_k(t) = 1$,
- (2) Moyenne $\sum_{k=0}^n k w_k(t) = nt$,
- (3) Variance $\sum_{k=0}^n (k - nt)^2 w_k(t) = nt(1-t)$.

On en tire des informations sur l'application de l'opérateur \mathcal{B}_n à des polynômes de degré 0, 1 et 2:

- (1) $f(x) = 1$: $(\mathcal{B}_n f)(t) = 1$,
- (2) $f(x) = x$: $(\mathcal{B}_n f)(t) = t$,
- (3) $f(x) = x^2$: $(\mathcal{B}_n f)(t) = t^2 + t(1-t)/n$,

Pour le dernier cas, on développe le carré de $k - nt$ dans l'identité de la variance, et on divise par n^2 .

Les polynômes de degré ≤ 1 sont donc reproduits exactement par \mathcal{B}_n , et on a pour le degré 2:

$$(\mathcal{B}_n(Ax^2 + Bx + C))(t) = At^2 + Bt + C + At(1-t)/n.$$

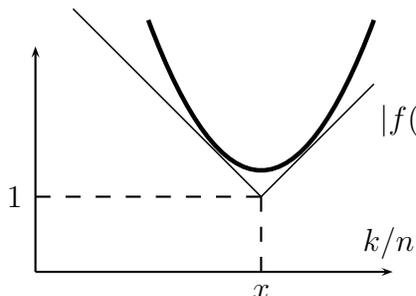
Ceci suffit à assurer la

Démonstration du théorème de Weierstrass par polynômes de Bernstein. Soit f continue sur $[0, 1]$, et $\epsilon > 0$. Montrons qu'il existe n tel que $\|\mathcal{B}_n f - f\|_\infty \leq \epsilon$.

Soit x un point de $[0, 1]$. Comme \mathcal{B}_n reproduit la constante $f(x)$,

$$(\mathcal{B}_n f)(x) - f(x) = \sum_0^n w_k(x)[f(k/n) - f(x)].$$

Introduisons le module de continuité de f :



$$|f(k/n) - f(x)| \leq \omega_f(|k/n - x|) \leq \left(1 + \frac{|k/n - x|}{h}\right) \omega_f(h),$$

par (48) (p. 62), où h est un réel > 0 non encore déterminé.

Soit $\xi = (k/n - x)/h$. Par $1 - 2|\xi| + \xi^2 \geq 0$,

$1 + \frac{|k/n - x|}{h} \leq \frac{3}{2} + \frac{(k/n - x)^2}{2h^2}$, où nous retrouvons la variance de la loi binomiale! Donc,

$$|(\mathcal{B}_n f)(x) - f(x)| \leq \omega_f(h) \left(\frac{3}{2} + \frac{x(1-x)}{2nh^2}\right), \text{ et enfin}$$

$$\|\mathcal{B}_n f - f\|_\infty \leq \text{const. } \omega_f(n^{-1/2}), \text{ en prenant } h = n^{-1/2}. \quad \square$$

On peut reprendre la preuve avec tout opérateur \mathcal{K}_n ayant les 3 propriétés: linéarité, positivité, et $f(t) = At^2 + Bt + C \Rightarrow \|\mathcal{K}_n f - f\|_\infty \leq |A|\varepsilon_n$, avec $\varepsilon_n \rightarrow 0$ quand $n \rightarrow \infty$ (Korovkin). Ici, $\varepsilon_n = 1/(4n)$.

Autres propriétés des polynômes de Bernstein.

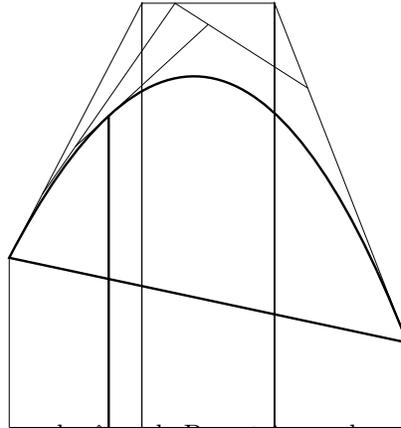
Quelques ref.: Z. Ditzian, V. Totik, *Moduli of Smoothness*, Springer, 1987. Zhongkai Li, Bernstein polynomials and modulus of continuity, *J. Approx. Theory* **102** (2000) 171-174. G.G. Lorentz, *Bernstein polynomials*, Univ. Toronto Press, 1953, 2^{ème} éd.: Chelsea, 1986; C.A. Micchelli, *Mathematical Aspects of Geometric Modeling*, SIAM (CBMS-NFS **65**) 1995; G. Farin, *Curves and Surfaces for Computer Aided Geometric Design. A Practical Guide*, Academic Press, 2nd ed., 1990.

La dérivée de $\mathcal{B}_n f$ est obtenue en appliquant l'opérateur \mathcal{B}_{n-1} aux différences divisées $n[f((k+1)/n) - f(k/n)]$, $k = 0, 1, \dots, n-1$.

$$\begin{aligned} \text{En effet, } \frac{\mathcal{B}_n f(x)}{dx} &= \sum_0^n \frac{n!}{k!(n-k)!} f\left(\frac{k}{n}\right) \frac{dx^k(1-x)^{n-k}}{dx} = \\ n \sum_1^n \frac{(n-1)!}{(k-1)!(n-k)!} f\left(\frac{k}{n}\right) x^{k-1}(1-x)^{n-k} &- n \sum_0^{n-1} \frac{(n-1)!}{k!(n-1-k)!} f\left(\frac{k}{n}\right) x^k(1-x)^{n-1-k} \\ &= \sum_0^{n-1} \frac{(n-1)!}{k!(n-1-k)!} \frac{f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right)}{1/n} x^k(1-x)^{n-1-k} \end{aligned}$$

La valeur de $\mathcal{B}_n f$ en x est identique à la valeur de \mathcal{B}_{n-1} aux combinaisons $xf((k+1)/n) + (1-x)f(k/n)$, $k = 0, 1, \dots, n-1$.

$$\begin{aligned} \text{En effet, } \mathcal{B}_n f(x) &= \sum_0^n \frac{n!}{k!(n-k)!} f\left(\frac{k}{n}\right) x^k(1-x)^{n-k} = \sum_0^n \left(\frac{(n-1)!}{(k-1)!(n-k)!} + \frac{(n-1)!}{k!(n-1-k)!}\right) f\left(\frac{k}{n}\right) x^k(1-x)^{n-k} = \\ \sum_0^{n-1} \frac{(n-1)!}{k!(n-1-k)!} \left[xf\left(\frac{k+1}{n}\right) + (1-x)f\left(\frac{k}{n}\right)\right] &x^k(1-x)^{n-1-k}. \end{aligned}$$



Construction d'une valeur d'un polynôme de Bernstein par la méthode de Casteljau Ici, $x = 1/4$.

Les approximations par polynômes de Bernstein ont d'importantes qualités d'ordre esthétique. On trouve des polynômes de Bernstein en carrosseries et coques (courbes et surfaces de Bézier), et en typographie informatique (voir § 7.8).

<http://www.worldscibooks.com/mathematics/7426.html>

Series on Concrete and Applicable Mathematics - Vol. 8

APPROXIMATION BY COMPLEX BERNSTEIN AND CONVOLUTION TYPE OPERATORS

by Sorin G Gal (University of Oradea, Romania)

World Scientific Publishing Co.

Table of Contents (128k) Preface (103k) Chapter 1: Bernstein-Type Operators of One Complex Variable (1,102k) http://www.worldscibooks.com/etextbook/7426/7426_chap01.pdf

352pp Pub. date: Aug 2009

Théorèmes de Jackson et Bernstein.

Pour mieux capturer l'erreur E_n de meilleure approximation sur $[a, b]$ par des polynômes de degré $\leq n$, on étudie des expressions de la forme $\int_a^b K_n(x, y) f(y) dy$, où K_n est un polynôme de degré $\leq n$ en x . Par exemple:

Théorème de Jackson. Si $f \in \mathcal{C}[-1, 1]$, $E_n \leq \text{const. } \omega_f(n^{-1})$.

En effet [Mha], on construit

$$p_n(x) = \frac{1}{\lambda_m} \int_{-\pi}^{\pi} \left[\frac{\sin((m+1/2)(\varphi-\theta))}{\sin((\varphi-\theta)/2)} \right]^4 F(\varphi) d\varphi,$$

où $F(\varphi) = f(\cos \varphi)$, m est la partie entière par défaut de $n/4$, et λ_n est tel que $p_n(x) \equiv 1$ si $f(x) \equiv 1$. On a bien $p_n \in \mathcal{P}_n$ puisque $\sin((m+1/2)(\varphi-\theta))/\sin((\varphi-\theta)/2) = \frac{e^{i(m+1/2)(\varphi-\theta)} - e^{i(m+1/2)(-\varphi+\theta)}}{e^{i(\varphi-\theta)/2} - e^{i(-\varphi+\theta)/2}} = 2 \sum_0^m \cos(k(\varphi-\theta))$

est un polynôme de degré $\leq m$ en $x = \cos \theta$ (augmenté d'une fonction impaire en φ qui disparaît dans l'intégrale). Ensuite, λ_m est l'intégrale de la quatrième puissance de la fonction ci-dessus, c'est-à-dire de $\sum_{-m}^m e^{ik(\varphi-\theta)}$, donc du carré de $\sum_{-2m}^{2m} (2m+1-|k|)e^{ik(\varphi-\theta)}$, ce qui fait 2π fois la somme des carrés des coefficients, soit $\lambda_m = 2\pi[2(1+4+\dots+4m^2) + (2m+1)^2] = 2\pi(2m+1)(8m^2+8m+3)/3 \sim \pi n^3/6$. Nous avons maintenant, pour θ fixé,

$$p_n(\cos \theta) - F(\theta) = \frac{1}{\lambda_m} \int_{-\pi}^{\pi} \left[\frac{\sin((m+1/2)\psi)}{\sin(\psi/2)} \right]^4 [F(\theta+\psi) - F(\theta)] d\psi,$$

où on a pris $\psi = \varphi - \theta$. On borne $|F(\theta+\psi) - F(\theta)|$ par $\omega_F(|\psi|) \leq (1+n|\psi|)\omega_F(n^{-1})$ par la propriété (48) de ω vue en p. 62. Finalement, on pose $\xi = (m+1/2)\psi$; le sinus du dénominateur est partout supérieur à $(2/\pi)|\psi|/2 = |\xi|/((m+1/2)\pi)$, et il reste une borne $(m+1/2)^2\pi^4(n+1)/\lambda_m$ fois $\omega_F(n^{-1})$ (qui est inférieur à

$\omega_f(n^{-1})$) fois $\int_{-\infty}^{\infty} |\xi|^{-3} \sin^4 \xi \, d\xi$, cette dernière intégrale étant convergente (c'est pour cela qu'on a dû prendre une quatrième puissance). □

Si $f \in \mathcal{C}^r$, on montre $E_n \leq c(r)n^{-r} \omega_{f^{(r)}}(n^{-1})$.

Si $f \in Lip_\alpha^{40}$, on a donc $E_n \leq \text{const. } n^{-\alpha}$. Réciproque (uniquement pour des polynômes trigonométriques):

Théorème de Bernstein. *Si $E_n(F) \leq \text{const. } n^{-\alpha}$, $0 < \alpha < 1$, alors $\omega_F(h) \leq \text{const. } h^\alpha$.*

En effet, soit P_n la meilleure approximation trigonométrique de degré $\leq n$ de F . Pour $h > 0$ fixé, et pour tout n entier, On a

$$\omega_F(h) \leq \omega_{F-P_n}(h) + \omega_{P_n}(h) \leq 2E_n + h\|P'_n\|_\infty \leq 2E_n + h[\|P'_n - P'_{n/2}\| + \|P'_{n/2} - P'_{n/4}\| + \dots].$$

Chaque $P_m - P_{m/2}$ est un polynôme trigonométrique de degré $\leq m$, donc, par une inégalité de Bernstein (cf. p. 43), la norme de la dérivée est bornée par le degré fois la norme de la fonction:

$$\|P'_m - P'_{m/2}\| \leq m\|P_m - P_{m/2}\| = m\|F - P_{m/2} - (F - P_m)\| \leq 2mE_{m/2}.$$

Et il reste

$$\omega_F(h) \leq 2E_n + h[2nE_{n/2} + nE_{n/4} + (n/2)E_{n/8} + \dots]$$

borné par const. fois $n^{-\alpha} + hn^{1-\alpha}[2^\alpha + 2^{-1+2\alpha} + 2^{-2+3\alpha} + \dots]$. La série converge. Il reste à choisir $n!$ On minimise en n $n^{-\alpha} + Chn^{1-\alpha}$, ce qui donne n de l'ordre de h^{-1} . □

7.5. Théorème de Stone-Weierstrass.

De nombreuses généralisations du théorème de Weierstrass ont abouti à cet énoncé remarquable:

Théorème de Stone-Weierstrass. *Soit I un espace compact, $\mathcal{C}(I)$ l'algèbre des applications continues de I dans \mathbb{C} , munie de la topologie de la convergence uniforme. Soit \mathcal{A} une sous-algèbre de $\mathcal{C}(I)$ telle que:*

- (1) \mathcal{A} sépare les points de I , c'est-à-dire $\forall x, y \in I, x \neq y: \exists f \in \mathcal{A} : f(x) \neq f(y)$,
- (2) $\forall x \in I, \exists f \in \mathcal{A} : f(x) \neq 0$,
- (3) $f \in \mathcal{A} \Rightarrow$ la fonction complexe conjuguée $\bar{f} \in \mathcal{A}$.

Alors, \mathcal{A} est dense dans $\mathcal{C}(I)$. ⁴¹

Remarquons que, par 3., on peut se limiter au cas réel: si $p \pm iq$ et $r \pm is \in \mathcal{A}$, p, q, pr et qs sont aussi dans \mathcal{A} : $p = [p + iq + (p - iq)]/2$, $pr = [(p + iq)(r + is) + (p - iq)(r - is) + (p - iq)(r + is) + (p + iq)(r - is)]/4$, etc.

Par 1. et 2., on peut toujours trouver un élément de \mathcal{A} prenant deux valeurs déterminées, soit A et B , en deux points distincts donnés de I : par 1., soit f_1 prenant deux valeurs distinctes en x' et x'' , l'une de ces deux valeurs, soit $f_1(x')$ étant donc non nulle, et, par 2. f_2 avec $f_2(x'') \neq 0$. On construit $f_3 = f_1^2 - f_1(x'')f_1$ qui est donc nulle en x'' et toujours non nulle en x' (où f_3 vaut $f_1(x')(f_1(x') - f_1(x'')) \neq 0$), et $f_4 = f_2 - cf_3$ avec c tel que $f_4(x') = 0$. Alors

$$A \frac{f_3}{f_3(x')} + B \frac{f_4}{f_4(x'')} \text{ convient.}$$

Une partie de la démonstration donnée dans le Cours d'analyse de J. Mawhin⁴² pp.708-711:

Si $g \in \overline{\mathcal{A}}$ avec $0 \leq g(x) \leq 1$ pour tout $x \in I$, alors $\sqrt{g} \in \overline{\mathcal{A}}$. En effet, \sqrt{g} peut être approché aussi bien que l'on veut par un polynôme en g (cf. approximation polynomiale de $\sqrt{\cdot}$, p. 61), ce polynôme en g pouvant lui-même être approché par un élément de \mathcal{A} .

⁴⁰Voir p. 121.

⁴¹Cf., par exemple, G. Choquet, *op.cit.*, chap. VI; A. Pinkus, Weierstrass and approximation theory, *J. Approx. Theory* **107** (2000) 1-66.

⁴²Analyse, *Fondements, techniques, évolution*, De Boeck Université, 1992

$f \in \overline{\mathcal{A}} \Rightarrow |f| \in \overline{\mathcal{A}}$. En effet, on prend $\|f\|_\infty \sqrt{\frac{f^2}{\|f\|_\infty^2}}$.

Si f et $g \in \overline{\mathcal{A}}$, $\max(f, g)$ et $\min(f, g) \in \overline{\mathcal{A}}$. En effet, \min et $\max(f, g) = (f + g \pm |f - g|)/2$.

Ensuite, pour f continue réelle sur I , on considère $P_{u,y} \in \mathcal{A}$ interpolant f en u et $y \in I$. Pour u fixé et x suffisamment près de y , $P_{u,y}(x) \geq f(x) - \varepsilon/2$. On recouvre le compact I par un nombre fini m de tels voisinages de façon à avoir $g_u := \max(P_{u,y_1}, \dots, P_{u,y_m}) \geq f - \varepsilon/2$ partout dans I . Dans un voisinage de u , on a aussi $g_u \leq f + \varepsilon/2$, et on recouvre maintenant I par des voisinages de u_1, \dots, u_r de sorte que $g := \min(g_{u_1}, \dots, g_{u_r}) \leq f + \varepsilon/2$ partout dans I . On a donc $g \in \overline{\mathcal{A}}$ vérifiant $\|g - f\|_\infty \leq \varepsilon/2$, et comme g est elle-même à moins d' $\varepsilon/2$ d'un élément p de \mathcal{A} , $\|f - p\|_\infty \leq \varepsilon$ \square

En pratique, on part d'une suite $\{\psi_k\}$ (éventuellement finie) de $\mathcal{C}(I)$, avec $\overline{\psi_k}$ dans la suite si ψ_k est dans la suite, et l'énoncé devient: si les ψ_k séparent les points de I (ils suffit que l'application identité soit dans \mathcal{A}), et si les ψ_k ne s'annulent pas tous en un même point de I (il suffit que \mathcal{A} contienne la fonction constante 1), toute fonction continue sur le compact I peut être approchée d'aussi près que l'on veut en $\|\cdot\|_\infty$ par des polynômes en les ψ_k .

Exemple. I est le cercle unité $\{z : |z| = 1\}$, $\{\psi_0, \psi_1\} = \{z, z^{-1} = \overline{z}\}$. Comme z et $1 = \psi_0\psi_1 \in \mathcal{A}$, les polynômes z^k , $k = \dots, -2, -1, 0, 1, 2, \dots$ forment une suite totale dans $\mathcal{C}(I)$, donc les **polynômes trigonométriques** $e^{ik\theta}$, $k = \dots, -2, -1, 0, 1, 2, \dots$ forment une suite totale dans l'espace des fonctions continues périodiques (c'est-à-dire $f(-\pi) = f(\pi)$) sur $[-\pi, \pi]$.

En réel: $\{\psi_0, \psi_1\} = \{\cos \theta, \sin \theta\}$ engendre les polynômes trigonométriques, et $1 = \cos^2 \theta + \sin^2 \theta$ est bien dans \mathcal{A} .

Cas des séries de Fourier de fonctions continues.

Ce qui précède ne veut pas dire que les séries de Fourier de fonctions continues périodiques convergent toujours uniformément! Soit \mathcal{T}_n l'espace engendré par $\exp(ik\theta)$, $k = -n, \dots, n$. Les sommes partielles de Fourier $S_n(\theta) = \sum_{-n}^n c_k \exp(ik\theta)$ sont les meilleures approximations possibles de f dans \mathcal{T}_n au sens de L^2 ; on a montré que, $\forall n$, il existe des coefficients $d_k^{(n)}$, $k = -n, \dots, n$, tels que $\|f - \sum_{-n}^n d_k^{(n)} \exp(ik\theta)\|_\infty \rightarrow 0$ quand $n \rightarrow \infty$, d'où on tire d'ailleurs également $\|f - \sum_{-n}^n d_k^{(n)} \exp(ik\theta)\|_2 \rightarrow 0$ et le caractère total de la suite $\{\exp(ik\theta)\}_{-\infty}^\infty$ dans $L^2(-\pi, \pi)$ ⁴³, mais les c_k ne sont pas optimaux dans \mathcal{C} .

Voir aussi Robert Feinerman, D. J. Newman: Completeness of $\{A \sin nx + B \cos nx\}$ on $[0, \pi]$, *Michigan Math. J.* **15**, nr . 3 (1968), 305-312.

On peut cependant construire des sommes de Fourier modifiées $\sigma_n := (S_0 + S_1 + \dots + S_{n-1})/n$ (**sommes de Fejér**), montrer que σ_n est l'application d'un opérateur linéaire **positif** à f et démontrer, un peu comme pour les polynômes de Bernstein, que $\|f - \sigma_n\|_\infty \rightarrow 0$ quand $n \rightarrow \infty$ si $f \in \mathcal{C}[-\pi, \pi]$ et $f(-\pi) = f(\pi)$. La vitesse de convergence des sommes de Fejér est cependant très médiocre (les normes d'erreur décroissent au mieux comme $1/n$), aussi

⁴³On peut montrer séparément que la suite $\{\exp(ik\theta)\}_{-\infty}^\infty$ est **maximale** dans $L^2(-\pi, \pi)$, cf. [Dav], pp.266-267: si $f \in L^2$ était orthogonale à tous les polynômes trigonométriques q , $\int_{-\pi}^\pi f(\theta)q(\theta)d\theta = 0 \Rightarrow \int_{-\pi}^\pi F(\theta)q'(\theta)d\theta = 0$ où F est la primitive de f s'annulant en $-\pi$, donc aussi en π [f ayant une moyenne nulle], donc $G = F -$ sa moyenne est une fonction **continue** orthogonale à tous les polynômes trigonométriques. Si G n'est pas la fonction nulle elle atteint une valeur $m \neq 0$, soit $m > 0$, et est $\geq m/2$ dans un intervalle $[\theta_0 - \delta, \theta_0 + \delta]$. Comme $1 - \cos \delta + \cos(\theta - \theta_0)$ est ≥ 1 dans le δ -voisinage fermé de θ_0 , et dans $[-\cos \delta, 1)$ hors de ce voisinage, $\int_{-\pi}^\pi G(\theta)[1 - \cos \delta + \cos(\theta - \theta_0)]^N d\theta > m/2 - \varepsilon$, impossible puisque doit être nul, donc G (et f) est la fonction nulle.

essaye-t-on de tirer parti du comportement des coefficients et des sommes de Fourier (p. 121). Voici encore un autre résultat classique de la littérature:

Théorème de Müntz. Si $0 = \lambda_0 < \lambda_1 < \dots, \lambda_n \rightarrow \infty$ quand $n \rightarrow \infty$, les puissances $x^{\lambda_0} = 1, x^{\lambda_1}, \dots$ forment une suite totale dans $\mathcal{C}[0, 1]$ si $\sum_{k=1}^{\infty} 1/\lambda_k = \infty$. Voir p.82

Pour des suites de **fonctions rationnelles**, cf.

Van Deun, J.; Bultheel, A.: Orthogonal rational functions and quadrature on an interval, *J. Comput. Appl. Math.* **153**, No.1-2, 487-495 (2003), où on trouve cet énoncé: la suite

$\left\{ \frac{x^n}{\prod_{k=1}^n (1 - x/a_k)} \right\}_0^{\infty}$, avec a_k réels et $1 < |a_k| \leq \infty$, est totale dans $\mathcal{C}[-1, 1]$ si et seulement

si $\sum_{k=1}^{\infty} (1 - |c_k|) = \infty$, où c_k est déterminé par $2a_k = c_k + 1/c_k$ et $|c_k| < 1$.

D. Figueiras, On the construction of a basis of $L_{\omega}^2(\mathbb{R})$ formed by pole-free rational functions, *L2omega.ps* étudie de façon très détaillée (nombreux exemples) des développements en fonctions orthogonales de la forme $p_n(x)/(1 + x^2)^{\lfloor n/2 \rfloor}$ sur \mathbb{R} ($p_n \in \mathcal{P}_n$).

7.6. Noyaux reproduisants, polynômes noyaux, représentants de Riesz.

Soit $\{b_0, \dots, b_n\}$ une base **orthonormale** de V , (58) se simplifie encore:

$$\hat{p} = Pf = \sum_{j=0}^{j=n} (f, b_j) b_j.$$

Dans la notation des physiciens (Dirac):

$$P = \sum_{j=0}^{j=n} |j\rangle \langle j|.$$

Si X est un espace de **fonctions**, muni du produit scalaire $(f, g) = \int_S f(x) \overline{g(x)} d\mu(x)$, on a aussi

$$\begin{aligned} (Pf)(y) &= \sum_{j=0}^{j=n} \left[\int_S f(x) \overline{b_j(x)} d\mu(x) \right] b_j(y) \\ &= \int_S f(x) \sum_{j=0}^{j=n} \overline{b_j(x)} b_j(y) d\mu(x) \\ &= (f, K_n(\cdot, y)), \end{aligned}$$

où $K_n(x, y) := \sum_{j=0}^n b_j(x) \overline{b_j(y)}$ est le **noyau reproduisant**⁴⁴ de V , en effet, $\forall f \in V, f(y) = (f, K_n(\cdot, y))$. Donc, $K_n(\cdot, y)$ est le **représentant de Riesz** de la forme $f \mapsto f(y)$.

Cette propriété reproduisante (sur V) est une voie d'accès à une approche de la convergence. Le comportement de $K_n(x, y)$ quand n est grand n'est cependant pas toujours de tout repos.

⁴⁴On dit aussi régénérateur (S.Colombo)

Ainsi, avec le produit scalaire $\int_{-1}^1 f(x)g(x)(1-x^2)^{-1/2}dx$ pour lequel on connaît $b_0 = \pi^{-1/2}$, $b_j(x) = (2/\pi)^{1/2}T_j(x)$ pour $j \geq 1$, on calcule

$$K_n(x, y) = \frac{2}{\pi} \sum_{j=0}^n T_j(x)T_j(y) = \frac{1}{2\pi} \left(\frac{\sin((n+1/2)(\theta_2 - \theta_1))}{\sin((\theta_2 - \theta_1)/2)} + \frac{\sin((n+1/2)(\theta_2 + \theta_1))}{\sin((\theta_2 + \theta_1)/2)} \right),$$

où $x = \cos \theta_1$, $y = \cos \theta_2$. Cette expression peut devenir très grande. D'ailleurs, avec des produits scalaires de forme $\int_a^b f(x)g(x)d\mu(x)$, il n'existe pas de fonction continue de deux variables K qui vérifierait $f(y) = (f, K(\cdot, y))$ pour toute fonction f continue sur $[a, b]$: on aurait $|f(y)| \leq \|f\| \|K(\cdot, y)\|$ (Cauchy-Schwarz), mais il existe des fonctions continues f de norme quadratique $\|f\|$ arbitrairement plus petite qu'une valeur ponctuelle $|f(y)|$: il faudrait $\|K(\cdot, y)\| = \infty$. L'"objet" qui peut alors prétendre servir de limite de $K_n(x, y)$ quand $n \rightarrow \infty$ ne peut être défini qu'au sens des **distributions** (distribution de Dirac $\delta(x - y)$).

Si V contient les constantes, on a aussi $C = (C, K_n(\cdot, y))$, donc, pour y fixé, $f(y) = (f(y), K_n(\cdot, y))$,

$$f(y) - (Pf)(y) = (f(y) - f(\cdot), K_n(\cdot, y)) = \int_S (f(y) - f(x)) \overline{K_n(x, y)} d\mu(x),$$

identité très utilisée en théorie de la convergence.

Il existe cependant des espaces à noyau reproduisant où la limite de $K_n(x, y)$ existe au sens classique: c'est le cas de l'espace des fonctions analytiques de carré de valeur absolue intégrable sur un domaine borné du plan complexe (espace de Bergman ⁴⁵).

Remarquons aussi que le noyau reproduisant se... reproduit lui-même: $K_n(y, z) = (K_n(\cdot, z), K_n(\cdot, y))$. En particulier, $K_n(y, y) = (K_n(\cdot, y), K_n(\cdot, y)) = \|K_n(\cdot, y)\|^2$.

Les noyaux reproduisants servent aussi à comparer l'approximation en moyenne quadratique et d'autres approximations. Ainsi, soit p une approximation de f dans V (meilleure approximation au sens d'une autre norme, interpolant, etc.), et q la meilleure approximation en moyenne quadratique, projection orthogonale de f dans V . On a

$$\begin{aligned} f - q &= f - p + p - q = f - p + (p - q, K_n) \quad \text{car } p - q \in V \\ &= f - p - (f - p, K_n) + (f - q, K_n) \\ &= f - p - (f - p, K_n) \quad \text{car } f - q \perp V. \end{aligned}$$

C'est ainsi que l'on a pu comparer les sommes partielles de développements en séries de polynômes de Tchebycheff et les meilleures approximations au sens de la norme du maximum (*constantes de Lebesgue*).

Dans le cas des **polynômes orthogonaux sur un intervalle réel**, on tire profit de la formule de Christoffel-Darboux (p. 100): le **polynôme noyau**

$$K_n(x, y) = \sum_{i=0}^n \varphi_i(x)\varphi_i(y) = \beta_{n+1} \frac{\varphi_{n+1}(x)\varphi_n(y) - \varphi_{n+1}(y)\varphi_n(x)}{x - y}$$

⁴⁵Cf. [Dav] pp. 319–322.

Ceci montre que, $\forall y$, K_n est un polynôme de degré $\leq n$ orthogonal à \mathcal{P}_{n-1} par rapport au produit scalaire associé à la mesure $(x-y)d\mu(x)$:

$$\forall p \in \mathcal{P}_{n-1}, \int_a^b p(x)K_n(x,y)(x-y)d\mu(x) = \int_a^b p(x)\beta_{n+1} \frac{\varphi_{n+1}(x)\varphi_n(y) - \varphi_{n+1}(y)\varphi_n(x)}{x-y} (x-y)d\mu(x) = 0.$$

Ce produit scalaire est **indéfini** si $y \in (a, b)$.

Une propriété extrême importante:

Théorème. Soit $y \notin (a, b)$ fixé. Parmi tous les polynômes $p \in \mathcal{P}_n$ vérifiant $p(y) = 1$, le

polynôme de norme $\left[\int_a^b p(x)^2 d\mu(x) \right]^{1/2}$ minimale est $p(x) = K_n(x, y)/K_n(y, y)$.

En effet, exprimons que p , avec $p(y) = 1$ est de norme minimale: tout autre polynôme valant également 1 en y est de la forme $p(x) + (x-y)q(x)$, avec $q \in \mathcal{P}_{n-1}$, donc,

$$\int_a^b (p(x) + (x-y)q(x))^2 d\mu(x) = \int_a^b p(x)^2 d\mu(x) + 2 \int_a^b p(x)q(x)(x-y)d\mu(x) + \int_a^b q(x)^2(x-y)^2 d\mu(x)$$

sera bien strictement supérieur au carré de la norme de p pour tout autre polynôme si p est orthogonal à \mathcal{P}_{n-1} par rapport à la mesure $(x-y)d\mu(x)$, donc si $p(x) = \text{constante } K_n(x, y)$, ce qui donne bien $p(x) = K_n(x, y)/K_n(y, y)$. \square

On appelle **fonction de Christoffel** associée à la mesure $d\mu$ la fonction $\lambda_n(x) = 1/K_{n-1}(x, x)$. Les constantes de Christoffel apparaissant dans la formule de quadrature de Gauss sont les valeurs de λ_n aux zéros de Φ_n .

Exercice. Soit w une fonction positive et continue sur (a, b) , et $\{\Phi_n(\cdot, \beta)\}$ les polynômes orthogonaux par rapport à la restriction de w sur (a, β) , avec $a < \beta \leq b$. Alors, chaque zéro $x_i(\beta)$ de $\Phi_n(\cdot, \beta)$ est une fonction croissante de β .

En effet⁴⁶, on a $\int_a^\beta \frac{\Phi_n^2(x, \beta)}{x - x_i(\beta)} w(x) dx = 0$. Dérivons en β :

$$\int_a^\beta \frac{2\Phi_n(x, \beta) \frac{\partial \Phi_n(x, \beta)}{\partial \beta}}{x - x_i(\beta)} w(x) dx + \frac{dx_i(\beta)}{d\beta} \int_a^\beta \frac{\Phi_n^2(x, \beta)}{(x - x_i(\beta))^2} w(x) dx + \frac{\Phi_n^2(\beta, \beta)}{\beta - x_i(\beta)} w(\beta) = 0.$$

Comme $\Phi_n(x) = \prod_1^n (x - x_k)$, $\frac{\partial \Phi_n(x)}{\partial \beta} = -\Phi_n(x) \sum_1^n x'_k (x - x_k)^{-1}$, la première intégrale se limite à

$$-2 \frac{dx_i(\beta)}{d\beta} \int_a^\beta \frac{\Phi_n^2(x, \beta)}{(x - x_i(\beta))^2} w(x) dx, \text{ et il reste } \frac{dx_i(\beta)}{d\beta} = \frac{\frac{\Phi_n^2(\beta, \beta)}{\beta - x_i(\beta)} w(\beta)}{\int_a^\beta \frac{\Phi_n^2(x, \beta)}{(x - x_i(\beta))^2} w(x) dx}.$$

7.7. Intervalles non bornés, problème des moments.

Cf. N.I. Akhiezer, *The Classical Moment Problem and some related Questions in Analysis*, Hafner, N.Y. 1965; Jean Dieudonné, *Éléments d'analyse*, vol. 2, Gauthier-Villars, Paris, 1968, exercices du § 15.13; P. Deift, *Orthogonal Polynomials: a Riemann-Hilbert Approach*, Courant Institute & Amer. Math. Soc., 1999, 2000,

⁴⁶Lemme 2 de D.K. Dimitrov: On a conjecture concerning monotonicity of zeros of ultraspherical polynomials, *J. Approx. Theory*, **85** (1996) 88-97.

§ 2.4: on dit que $d\mu$ correspond à un problème des moments déterminé si la suite numérique $\{\mu_k\}_{k=0}^\infty$ des moments suffit à déterminer μ , c'est-à-dire que les seules fonctions croissantes bornées ν vérifiant

$$\int_{\mathbb{R}} x^k d\nu(x) = \mu_k, \quad k = 0, 1, \dots$$

sont $\nu(x) = \mu(x) + \text{constante}$, à une infinité dénombrable de points de discontinuité près.

La famille de mesures $\{\mu\}$ admettant une suite donnée de moments admet aussi la même suite de polynômes orthogonaux $\{\Phi_n\}$, et donc aussi les mêmes coefficients de récurrence $\{\alpha_n, \beta_n\}$, et encore les mêmes noeuds x_k et poids H_k des formules d'intégration de Gauss pour tous les degrés.

Relations entre convergence, totalité de la suite $\{\Phi_n\}$ et problème des moments:

- (1) L'erreur de formule de Gauss peut s'écrire

$$\int_{\mathbb{R}} f(t) d\mu(t) - \sum_1^n H_k f(x_k) = \int_{\mathbb{R}} [f(t) - p_{\text{interp}}(t)] d\mu(t),$$

où p_{interp} interpole f aux points x_k . On s'attend à ce que le membre de droite soit petit si les polynômes forment une partie dense de $L^2_{d\mu}$; d'autre part, $\sum H_k f(x_k)$ "ne sait pas" vers quoi elle doit converger si plusieurs mesures μ peuvent figurer dans le membre de gauche. Tout ceci donne des idées, mais encore rien de rigoureux.

- (2) D'ailleurs, on peut utiliser la formule de Gauss pour estimer μ , en intégrant une fonction échelon:

$$\sum_{x_k < x^*} H_k \approx? \int_{-\infty}^{x^*} d\mu(t) = \mu(x^*) - \mu(-\infty).$$

Une approximation plus lisse s'obtient en prenant la partie imaginaire d'un logarithme⁴⁷:

$$\mu(x^*) - \mu(-\infty) = \lim_{\substack{y^* \rightarrow 0 \\ y^* > 0}} \frac{-1}{\pi} \text{Im} \int_{\mathbb{R}} \text{Log}(t - z) d\mu(t) \approx? \frac{-1}{\pi} \text{Im} \sum H_k \text{Log}(x_k - z),$$

où $z = x^* + iy^*$, et où Log est la détermination principale du logarithme ($\text{Im}(\text{Log } x) = 0$ si $x > 0$).

- (3) Ceci nous amène (dérivation en x^*) à regarder de plus près la formule de Gauss appliquée à l'intégration de $(z - t)^{-1}$, pour z non réel donné:

$$F_n(z) := \sum_1^n \frac{H_k}{z - x_k} = \frac{\psi_n(z)}{\varphi_n(z)} \approx? F(z) := \int_{\mathbb{R}} \frac{d\mu(t)}{z - t},$$

d'après des formules vues à la section § 3.7, p. 98.

Les coefficients $c_k(z)$ du développement de $(z - t)^{-1}$ vérifient au moins l'inégalité de Bessel $\sum_0^\infty |c_k(z)|^2 \leq \int_{\mathbb{R}} |z - t|^{-2} d\mu(t)$, et on peut distinguer ce qui dépend de $F(z)$, qui est inconnu, et ce qui ne dépend que de la

suite connue des moments, donc des suites $\{\varphi_n\}$ et $\{\psi_n\}$: $c_k(z) = \int_{\mathbb{R}} \frac{\varphi_k(t)}{z - t} d\mu(t)$ vérifie la même récurrence $\beta_{k+1}c_{k+1}(z) = (z - \alpha_k)c_k(z) - \beta_k c_{k-1}(z)$ que φ_k à partir de $k = 1$. En tenant compte de $\beta_1 c_1(z) = (z - \alpha_0)c_0(z) - \sqrt{\mu_0}$, on trouve $c_k(z) = F(z)\varphi_k(z) - \psi_k(z)$, et reprenons Bessel:

$$\sum_{k=0}^\infty |F(z)\varphi_k(z) - \psi_k(z)|^2 \leq \int_{\mathbb{R}} \frac{d\mu(t)}{(z - t)(\bar{z} - t)} = \frac{F(z) - \overline{F(z)}}{z - \bar{z}}, \quad (81)$$

ce qui est une relation de la forme $(F(z) - C(z))(\overline{F(z)} - \overline{C(z)}) \leq R^2(z)$ pour le nombre complexe $F(z)$ qui doit donc être contenu dans un **disque** $D(z)$ parfaitement calculable (bien qu'un peu compliqué...) à partir de la suite des moments et de ce qui en dépend.

Voici une première proposition rigoureuse: *Si les polynômes forment une partie dense de $L^2_{d\mu}$, $F(z)$ est situé sur la frontière du disque $D(z)$, pour tout z non réel.* En effet, l'inégalité de Bessel devient l'égalité de Parseval. \square

On démontre aussi la **réciproque**: *si $F(z)$ se trouve sur la frontière de $D(z)$ pour un z non réel, il en est ainsi pour tous les z non réels, et les polynômes forment une partie dense de $L^2_{d\mu}$.* La démonstration établit le caractère essentiellement autoadjoint de l'opérateur \mathbf{T} limite des opérateurs \mathbf{T}_n vus en (67), et on examine la représentation spectrale de l'opérateur résolvant $(z\mathbf{I} - \mathbf{T})^{-1}$.

Tout se simplifie si le disque $D(z)$ est réduit à un point (cas **déterminé** du problème des moments). C'est certainement le cas si $\sum_0^\infty |\varphi_k(z)|^2 = \infty$: en effet, si on remplace F par $G \neq F$ dans (81), le membre de

⁴⁷L'égalité de la limite $y^* \rightarrow 0$ est la formule de Stieltjes-Perron.

gauche est le carré de $\|G\Phi - \Psi\| \geq |G - F| \|\Phi\| - \|F\Phi - \Psi\| = \infty$ ne peut plus être dans n'importe quel disque borné.

Des conditions suffisantes assurant que $d\mu$ correspond à un problème des moments déterminé sont:

$$\exists \beta > 0 : \int_{-\infty}^{\infty} e^{\beta|x|} d\mu(x) < \infty; \quad \sum_{n=1}^{\infty} (\mu_{2n})^{-1/(2n)} = \infty \text{ (Carleman).}$$

La suite de polynômes de Laguerre $\{L_n^\alpha\}$ est donc totale dans $L_{d\mu}^2$ pour $\mu : \mu(x) = 0$ si $x \leq 0$; $\mu(x) = \int_0^x t^\alpha e^{-t} dt$ si $x \geq 0$. ($\alpha > -1$).

La suite de polynômes d'Hermite $\{H_n\}$ est donc totale dans $L_{d\mu}^2$ pour $\mu : \mu(x) = \int_{-\infty}^x e^{-t^2} dt$.

7.8. Arcs de Bézier en typographie informatique.

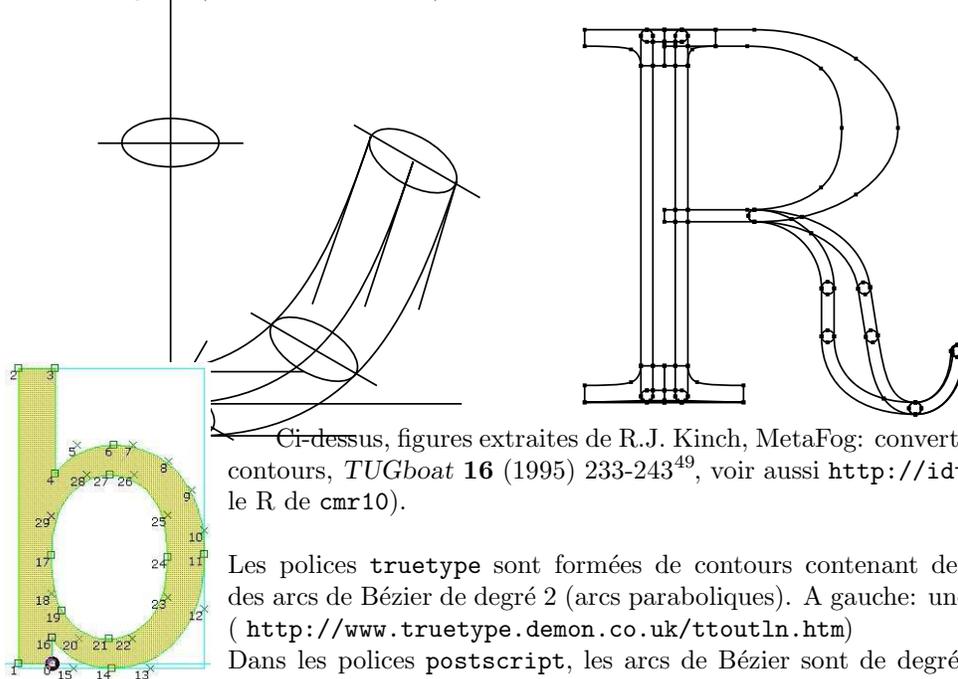
Un *arc de Bézier*⁴⁸ est décrit par deux polynômes de Bernstein

$$x = x(t) = \sum_{k=0}^n \xi_k \binom{n}{k} t^k (1-t)^{n-k},$$

$$y = y(t) = \sum_{k=0}^n \eta_k \binom{n}{k} t^k (1-t)^{n-k}.$$

Les $n + 1$ points (ξ_k, η_k) sont appelés *points de contrôle* de l'arc.

Le programme METAFONT de D.E. Knuth (cf. <http://www-cs-faculty.stanford.edu/~knuth/> voir aussi <http://www.loria.fr/tex/fontes.html>). décrit un caractère comme une armature formée d'arcs de Bézier de degré 3 (splines sous tension) parcourue par un pinceau, plume, feutre ou autre calame



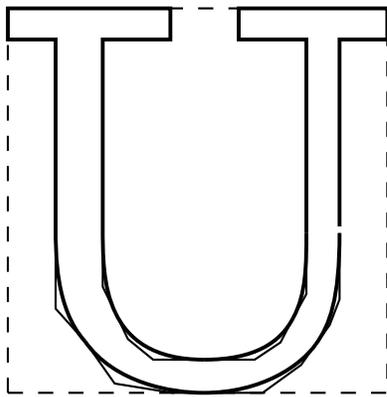
Ci-dessus, figures extraites de R.J. Kinch, MetaFog: converting METAFONT shapes to contours, *TUGboat* 16 (1995) 233-243⁴⁹, voir aussi <http://idt.net/~kinch/> (à droite: le R de cmr10).

Les polices **truetype** sont formées de contours contenant des segments rectilignes et des arcs de Bézier de degré 2 (arcs paraboliques). A gauche: une lettre b monotype arial. (<http://www.truetype.demon.co.uk/ttoutln.htm>)

Dans les polices **postscript**, les arcs de Bézier sont de degré 3. Ainsi, la lettre U de Courier (à gauche) contient 8 arcs de Bézier

⁴⁸Pierre Bézier, (1910-1999), a développé la génération numérique de courbes et de surfaces chez le constructeur automobile R... Paul de Casteljau, lui, travaillait chez Citroën (Un inconnu célèbre, Pierre Bézier, *Science & Vie Micro* n° 69, fév. 1990 [documentation aimablement communiquée par D. Grolaux]).

⁴⁹Un grand merci à M. Pierre Bulens qui a communiqué cette documentation.



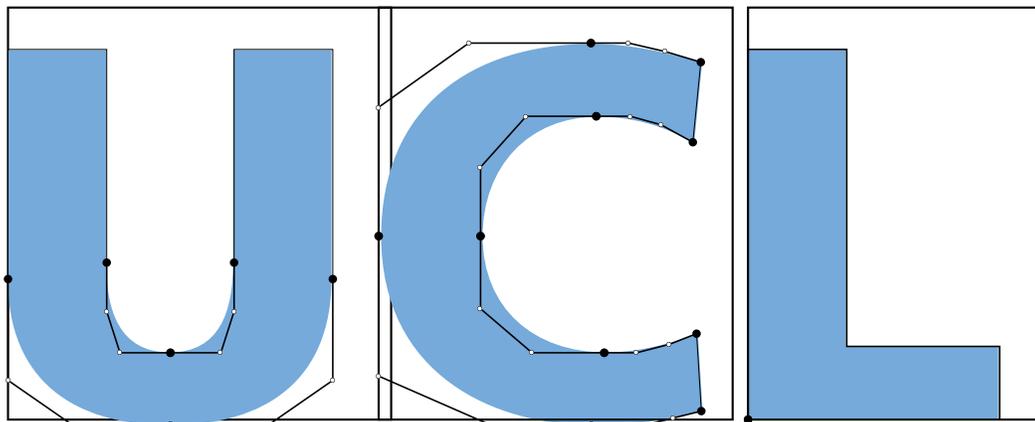
```

newpath 0 580 moveto 0 533 lineto 72 533 lineto
72 251 lineto
72 209 72 127 116 78 curveto
161 14 248 0 295 0 curveto
384 0 442 42 464 73 curveto
485 102 500 141 500 241 curveto
500 533 lineto 572 533 lineto 572 580 lineto
348 580 lineto 348 533 lineto 450 533 lineto
450 241 lineto
450 196 450 149 430 113 curveto
411 76 372 50 296 50 curveto
219 50 181 81 162 121 curveto
143 160 143 210 143 251 curveto
143 533 lineto 245 533 lineto 245 580 lineto
closepath
    
```

Ce caractère est inscrit dans un rectangle de 572×580 unités. 'xy moveto' signifie: placer le point courant en (x, y), 'xy lineto': tracer le segment rectiligne jusque (x, y), 'x₂y₂x₃y₃x₄y₄ curveto': tracer l'arc de Bézier de degré 3 partant du point courant à (x₄, y₄), avec (x₂, y₂) et (x₃, y₃) comme points de contrôle intermédiaires.

L'UCL a adopté pour ses communications officielles⁵⁰ la police **Frutiger** dont voici un échantillon

Frutiger-Black:



Les contours sont:

U:

```

612 698 moveto
426 698 lineto
426 296 lineto
426 204 401 126 306 126 curveto
211 126 186 204 186 296 curveto
186 698 lineto
0 698 lineto
0 265 lineto
0 74 124 -12 306 -12 curveto
488 -12 612 74 612 265 curveto
closepath
    
```

C:

```

599 162 moveto
548 142 486 126 425 126 curveto
289 126 192 209 192 346 curveto
192 476 278 572 410 572 curveto
474 572 532 556 592 523 curveto
607 674 lineto
540 694 471 710 400 710 curveto
170 710 0 589 0 346 curveto
0 82 215 -12 401 -12 curveto
496 -12 555 3 608 16 curveto
closepath
    
```

L:

```

0 0 moveto
474 138 lineto
474 0 lineto
186 138 lineto
186 698 lineto
0 698 lineto
closepath
    
```

Sur les B-splines et les NURBS (Non Uniform Rational B-Splines), voir *An Interactive Introduction to Splines* <http://www.ibiblio.org/e-notes/Splines/Intro.htm>

Voir aussi la belle page de Luc Devroye <http://cgm.cs.mcgill.ca/~luc/bezier.html>

CHAPITRE 4

Interpolation et applications.

Interpolation, intégration.

1. Interpolation.

Interpoler: *Introduire dans un ouvrage des passages qui ne sont pas dans l'original.
Math. Assigner à une quantité une valeur intermédiaire entre deux valeurs
directement calculées ou observées.*

Larousse.

Interpoler une fonction f consiste à “faire passer” le graphe d’une fonction p appartenant à un espace V de dimension finie par des points $(x_i, f(x_i))$ donnés en nombre égal à la dimension de V . L’estimation de la qualité de l’approximation de f par p est entièrement concentrée aux points x_i , on n’a besoin de ne rien savoir d’autre que les valeurs $f(x_i)$ pour déterminer p . Il peut sembler curieux de voir apparaître si tard le thème de l’interpolation dans

ce cours d’analyse numérique, alors que l’interpolation est le plus vieux procédé numérique connu (nombreux témoignages remontant à la plus haute antiquité). Bien sûr, l’interpolation reste l’outil classique par excellence, mais les progrès de l’analyse numérique ont subordonné cet outil à d’autres objectifs (approximation, lissage, filtrage, etc.)

A partir du moment où on n’est plus tenu de gérer des tables rigides (cadre de l’épure obligé jusqu’au ... 20^{ème} siècle bien avancé), et faute d’être submergé par le nombre de degrés de liberté devenus disponibles: où interpoler, quels points utiliser, il a fallu domestiquer cette abondance de moyens en préparant la voie par d’autres techniques.

Ainsi, on a rencontré des effets d’interpolation non sollicités. De par les propriétés d’oscillation qui la caractérise, la meilleure approximation polynomiale de degré $\leq n$ au sens de Tchebycheff interpole la fonction à approcher en au moins $n + 1$ points.

- (2) Les meilleures approximations au sens de la norme $\| \cdot \|_1$ se caractérisent par des conditions d’interpolation très explicites, (voir § 5, p. 118).
- (3) Par orthogonalité, on a vu que les erreurs de meilleures approximations dans des espaces préhilbertiens doivent changer de signe un nombre convenable de fois: l’approximation a donc des propriétés d’interpolation.

Développons ce dernier cas, et montrons comment on passe de l’approximation en moyenne à l’interpolation proprement dite:

On dispose de fonctions indépendantes $\varphi_0, \dots, \varphi_n$, et voyons si on peut trouver une combinaison $\sum_{k=0}^n a_k \varphi_k$ prenant des valeurs données y_0, \dots, y_n en des points x_0, \dots, x_n donnés. Réponse: orthogonaliser les fonctions $\varphi_0, \dots, \varphi_n$ en Φ_0, \dots, Φ_n selon un produit scalaire discret $(f, g) = \sum_{j=0}^n w_j f(x_j) \overline{g(x_j)}$ et construire les approximations successives $S_i = \sum_{k=0}^i c_k \Phi_k$, avec $c_k = (y, \Phi_k) / \|\Phi_k\|^2$.

Les S_i sont des approximations au sens des moindres carrés de y et on peut se contenter de $i < n$. Cependant, S_n interpole y en les $n + 1$ points x_0, \dots, x_n : $\|y - S_n\|$ doit être le plus petit possible et est nul si S_n interpole y . Encore faut-il vérifier que l'on peut interpolier n'importe quel ensemble de $n + 1$ données par une combinaison de $\varphi_0, \dots, \varphi_n$.

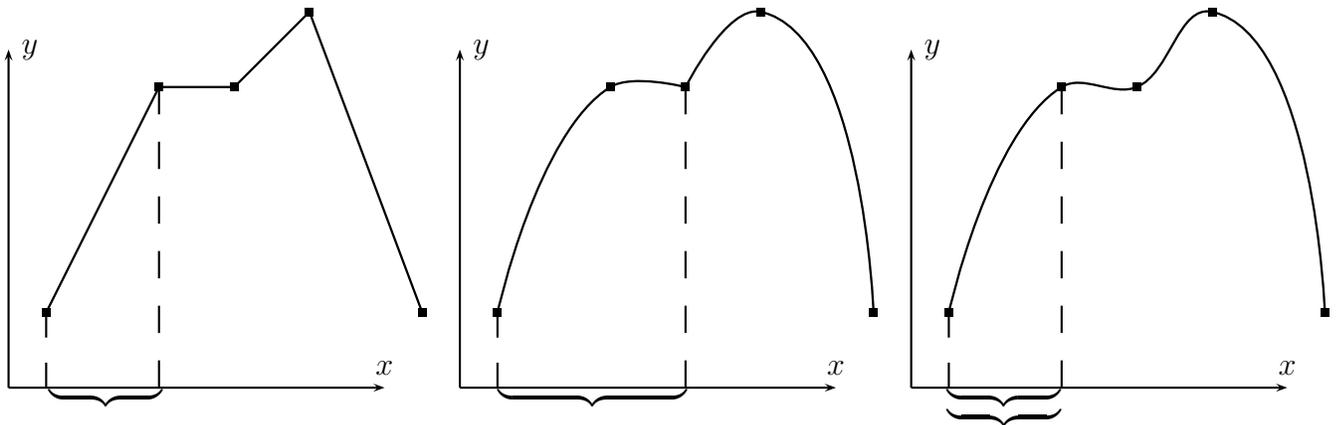
La fonction matlab `polyfit(x,y,n)` donne l'interpolant polynomial aux points (x_i, y_i) si $n + 1 =$ le nombre de ces points.

1.1. Interpolation polynomiale classique.

L'interpolation traditionnelle de f dans $[a, b]$ consiste à y approcher f par la fonction du premier degré $\alpha x + \beta$ prenant les valeurs $f(a)$ en a et $f(b)$ en b . Problème élémentaire résolu par

$$p(x) = \frac{[f(b) - f(a)]x + bf(a) - af(b)}{b - a} = f(a)\frac{b - x}{b - a} + f(b)\frac{x - a}{b - a} = f(a) + \frac{f(b) - f(a)}{b - a} (x - a). \tag{82}$$

On remédie à la performance souvent médiocre de cet interpolant, soit en réduisant l'intervalle $[a, b]$ et en reprenant l'interpolation soit en augmentant le degré de p , soit encore en modifiant les conditions d'interpolation. Ainsi, dans la figure ci-dessous, on a d'abord interpolé linéairement entre deux points consécutifs, puis interpolé par du second degré en *trois* valeurs consécutives, enfin par des polynômes du troisième degré entre deux points consécutifs, mais en s'arrangeant pour avoir une dérivée continue (interpolation au sens d'Hermite).



Déterminons $p \in \mathcal{P}_n$, en supposant f disponible aux points x_0, \dots, x_n *distincts*:

$$p(x_i) = \alpha x_i^n + \alpha' x_i^{n-1} + \dots + \alpha^{(n)} = f(x_i), \quad i = 0, 1, \dots, n \tag{83}$$

ce qui représente $n + 1$ équations linéaires pour les $n + 1$ coefficients $\alpha, \alpha', \dots, \alpha^{(n)}$ de $p \in \mathcal{P}_n$. La solubilité de ce système d'équations s'établit en principe par l'étude de son déterminant

$$\begin{vmatrix} x_0^n & x_0^{n-1} & \dots & x_0 & 1 \\ x_1^n & x_1^{n-1} & \dots & x_1 & 1 \\ \dots & \dots & \dots & \dots & \dots \\ x_n^n & x_n^{n-1} & \dots & x_n & 1 \end{vmatrix}, \text{ ce qui peut d'ailleurs être fait: il s'agit du } \mathbf{d\acute{e}terminant de Van-}$$

dermonde qui vaut $\prod_{i=0}^{n-1} \prod_{j=i+1}^n (x_i - x_j)$, effectivement non nul dès que les x_i sont distincts.

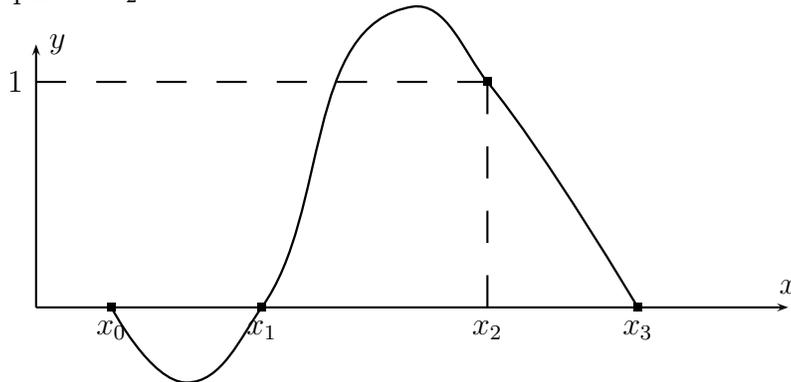
Il est beaucoup plus élégant et rapide de raisonner comme suit: le déterminant du système d'équations $\{p(x_i) = f(x_i)\}, i = 0, \dots, n$ est non nul si et seulement si le polynôme nul de \mathcal{P}_n est le *seul* à vérifier les équations homogènes: $p(x_0) = p(x_1) = \dots = p(x_n) = 0$ n'est possible que si p a $n + 1$ zéros distincts $\Rightarrow p$ est de degré $> n$ ou doit être le polynôme nul. On n'a même pas dû considérer le détail des équations de (83).

Les formulations usuelles de la solution du problème d'interpolation classique évitent également l'écriture trop rigide de (83). On adoptera en fait des bases de \mathcal{P}_n automatiquement bien adaptées au problème.

Formulation de Lagrange. Résolvons le problème d'interpolation particulier suivant: soit ℓ_j le polynôme de \mathcal{P}_n qui s'annule en tous les x_i sauf x_j , où il vaut 1: $\ell_j(x_i) = \delta_{i,j}, i = 0, 1, \dots, n$. Comme ℓ_j doit s'annuler en $x_0, x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n$, il doit admettre la factorisation $\ell_j(x) = K(x - x_0)(x - x_1) \dots (x - x_{j-1})(x - x_{j+1}) \dots (x - x_n)$, où K est une constante puisque la factorisation contient déjà n facteurs du premier degré. Cette constante est telle que $\ell_j(x_j) = 1$, ce qui détermine entièrement le polynôme $\ell_j \in \mathcal{P}_n$:

$$\ell_j(x) = \prod_{0 \leq k \leq n, k \neq j} \frac{x - x_k}{x_j - x_k}. \quad (84)$$

Graphe de ℓ_2 :



La solution du problème d'interpolation dans \mathcal{P}_n est alors

$$p = \sum_{i=0}^n f(x_i) \ell_i. \quad (85)$$

En effet, $p \in \mathcal{P}_n$ puisque p est une combinaison linéaire des ℓ_i ; $p(x_j) = \sum_i f(x_i) \ell_i(x_j) = \sum_i f(x_i) \delta_{i,j} = f(x_j)$.

Le grand avantage de la forme (85) est de mettre en évidence le rôle des valeurs de f . On voit par exemple que l'interpolant est le résultat d'une projection linéaire sur \mathcal{P}_n .

Exercice: vérifier $p(x) \equiv \sum_{i=0}^n p(x_i) \ell_i(x), \forall p \in \mathcal{P}_n$ pour de faibles valeurs de n .

Exercice: Montrez que le coefficient de x^n de l'interpolant de degré $\leq n$ de f en x_0, x_1, \dots, x_n distincts est la valeur de la forme linéaire

$$a_{n,\text{interp}_n}(f) = \sum_{i=0}^n \frac{f(x_i)}{\prod_{\substack{0 \leq k \leq n \\ k \neq i}} (x_i - x_k)} \quad (86)$$

En particulier, (86) donne donc le coefficient de x^n d'un élément de \mathcal{P}_n à partir de $n+1$ valeurs ponctuelles. C'est utilisé dans l'étude de propriétés extrémales de polynômes (chap. 2).

Interpolation cubique d'Hermite. On cherche le polynôme de \mathcal{P}_3 qui prend deux valeurs imposées y_0 et y_1 en deux points donnés x_0 et x_1 , et deux valeurs imposées de la *dérivée* y_2 et y_3 aux deux *mêmes* points:

$$p(x_0) = y_0, \quad p(x_1) = y_1; \quad p'(x_0) = y_2, \quad p'(x_1) = y_3.$$

Comme il y a quatre coefficients à déterminer dans $p(x) = a + bx + cx^2 + dx^3$, on écrit les équations $a + bx_0 + cx_0^2 + dx_0^3 = y_0$, $a + bx_1 + cx_1^2 + dx_1^3 = y_1$, $b + 2cx_0 + 3dx_0^2 = y_2$, $b + 2cx_1 + 3dx_1^2 = y_3$.

C'est un peu pénible, mais le déterminant se factorise joliment (tiens, tiens) en $-(x_1 - x_0)^4$, et on a

$$\begin{aligned} a &= \frac{x_0^2(3x_0 - x_1)y_0 - x_1^2(3x_1 - x_0)y_1}{(x_0 - x_1)^3} - \frac{x_0x_1(x_1y_2 + x_0y_3)}{(x_0 - x_1)^2}, \\ b &= \frac{6x_0x_1(y_0 - y_1)}{(x_1 - x_0)^3} + \frac{x_1(2x_0 + x_1)y_2 + x_0(2x_1 + x_0)y_3}{(x_1 - x_0)^2}, \\ c &= \frac{3(x_0 + x_1)(y_0 - y_1)}{(x_0 - x_1)^3} - \frac{(x_0 + 2x_1)y_2 + (2x_0 + x_1)y_3}{(x_0 - x_1)^2}, \\ d &= \frac{2(y_0 - y_1)}{(x_1 - x_0)^3} + \frac{y_2 + y_3}{(x_0 - x_1)^2}. \end{aligned}$$

Ouf! On cherchera une façon un peu plus systématique de

- (1) Poser le problème: on précisera comment rechercher un élément d'un espace vectoriel à partir de valeurs prises par des formes données.
- (2) Étudier l'existence et l'unicité de la solution: on se ramènera à un système d'équations linéaires.
- (3) Représenter la solution à partir de la *base duale* (ou *biorthogonale*) aux formes données.

1.2. Interpolation: cadre général.

Soit $\varphi_0, \varphi_1, \dots$ des éléments indépendants d'un espace vectoriel X , $\lambda_0, \lambda_1, \dots$ des formes linéaires indépendantes sur X ; V_n le sous-espace de X engendré par $\varphi_0, \dots, \varphi_n$, V_n^* le sous-espace du dual de X engendré par $\lambda_0, \dots, \lambda_n$.

Interpoler une suite de scalaires $\{y_0, y_1, \dots, y_n\}$ par un élément de V_n sur $\lambda_0, \dots, \lambda_n$ consiste à trouver $\varphi \in V_n$ (**l'interpolant**) tel que $\lambda_i(\varphi) = y_i$, $i = 0, 1, \dots, n$ (**problème d'interpolation sur $V_n \times V_n^*$**).

Posons $\varphi = \sum_{j=0}^n x_j \varphi_j$, le problème revient à trouver x_0, \dots, x_n tels que

$$\sum_{j=0}^n \lambda_i(\varphi_j) x_j = y_i \quad i = 0, \dots, n.$$

Système d'équations linéaires, donc, immédiatement

Proposition. *La solution du problème d'interpolation sur $V_n \times V_n^*$ existe et est unique si $\det[\lambda_i(\varphi_j)]_{i,j=0}^n \neq 0$.*

On dit alors que $\{\varphi_0, \dots, \varphi_n\}$ est *unisolvant* sur $\{\lambda_0, \dots, \lambda_n\}$.

Comme on a

$$\begin{bmatrix} \lambda_0(\varphi_0) & \cdots & \lambda_0(\varphi_n) & y_0 \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_n(\varphi_0) & \cdots & \lambda_n(\varphi_n) & y_n \\ \varphi_0 & \cdots & \varphi_n & \varphi \end{bmatrix} \begin{bmatrix} x_0 \\ \cdot \\ x_n \\ -1 \end{bmatrix} = 0,$$

$$\begin{vmatrix} \lambda_0(\varphi_0) & \cdots & \lambda_0(\varphi_n) & y_0 \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_n(\varphi_0) & \cdots & \lambda_n(\varphi_n) & y_n \\ \varphi_0 & \cdots & \varphi_n & \varphi \end{vmatrix} = 0,$$

l'interpolant est

$$\varphi = - \frac{\begin{vmatrix} \lambda_0(\varphi_0) & \cdots & \lambda_0(\varphi_n) & y_0 \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_n(\varphi_0) & \cdots & \lambda_n(\varphi_n) & y_n \\ \varphi_0 & \cdots & \varphi_n & 0 \end{vmatrix}}{\begin{vmatrix} \lambda_0(\varphi_0) & \cdots & \lambda_0(\varphi_n) \\ \cdots & \cdots & \cdots \\ \lambda_n(\varphi_0) & \cdots & \lambda_n(\varphi_n) \end{vmatrix}}.$$

On évite des calculs de déterminants en tirant parti des équivalences suivantes:

Proposition. *Les quatre énoncés suivants sont équivalents:*

- (1) $\{\varphi_0, \dots, \varphi_n\}$ est unisolvant sur $\{\lambda_0, \dots, \lambda_n\}$. Pour tout $[y_0, \dots, y_n] \in \mathbb{R}^{n+1}$ ou \mathbb{C}^{n+1} , il existe donc toujours exactement une combinaison linéaire φ de $\varphi_0, \dots, \varphi_n$ vérifiant $\lambda_i(\varphi) = y_i, i = 0, \dots, n$.
- (2) $\det[\lambda_i(\varphi_j)]_{i,j=0}^n \neq 0$.
- (3) Seul l'élément nul de V_n interpole la suite nulle $\{0, \dots, 0\}$ sur $\lambda_0, \dots, \lambda_n$.
- (4) On peut trouver $n + 1$ éléments ℓ_0, \dots, ℓ_n de V_n vérifiant $\lambda_i(\ell_j) = \delta_{i,j}, i, j = 0, \dots, n$ (**base duale**, ou **biorthogonale**, ou encore **base de Lagrange** relativement aux formes $\lambda_0, \dots, \lambda_n$).

Démonstration.

On a déjà vu (1) \iff (2). (2) \iff (3): si le déterminant est non nul, le système de $n + 1$ équations homogènes à $n + 1$ inconnues n'admet que la solution nulle; et si les équations homogènes n'admettent pas de solution non triviale, le rang de la matrice carrée est maximal \implies déterminant non nul.

(1) ou (2) \implies (4): comme les systèmes d'équations sont solubles pour tout second membre, on choisit pour second membre la $i^{\text{ème}}$ colonne de la matrice identité, et la solution fournit les coefficients du développement de ℓ_i dans la base $\{\varphi_0, \dots, \varphi_n\}$.

(4) \Rightarrow (1): (c'est assez subtil). Pour un second membre quelconque $[y_0, \dots, y_n]$, formons la combinaison linéaire

$$\boxed{\varphi = \sum_{j=0}^n y_j \ell_j.} \quad (87)$$

Constatons que chaque $\lambda_i(\varphi)$ vaut bien y_i :

$$\lambda_i(\varphi) = \lambda_i \left(\sum_{j=0}^n y_j \ell_j \right) = \sum_{j=0}^n y_j \lambda_i(\ell_j) = \sum_{j=0}^n y_j \delta_{i,j} = y_i, \text{ pour } i = 0, \dots, n. \quad \square$$

Le point (3) signifie le caractère **injectif** de l'opérateur représenté par la matrice des $\lambda_i(\varphi_j)$; le point (4) signifie le caractère **surjectif** de cet opérateur (remarque faite par un étudiant).

Exemples illustrant le point (3).

(1) **Interpolation polynomiale classique.**

Soit $V_n = \mathcal{P}_n, x_0, \dots, x_n$ des points *distincts* de \mathbb{R} ou \mathbb{C} , et les formes $\lambda_i(\varphi) := \varphi(x_i)$, $i = 0, \dots, n$.

Avec $\varphi_j = x^j$, on peut vérifier directement que $\det[x_i^j]_{i,j=0}^n \neq 0$ (déterminant de Vandermonde), ou, plus rapidement, qu'un polynôme non nul de degré $\leq n$ ne peut s'annuler aux $n + 1$ points x_0, \dots, x_n .

On retrouve évidemment le cas élémentaire.

(2) **Interpolation polynomiale d'Hermite, ou confluente.**

Soit $V_n = \mathcal{P}_n, x_0, \dots, x_\ell$ des points distincts, $\ell \leq n$, auxquels on associe une ou plusieurs formes $\varphi(x_i), \varphi'(x_i), \dots$ faisant appel à des dérivées successives, de sorte que le nombre total de formes utilisées soit bien $n + 1$: si les formes relatives à x_i sont $\varphi(x_i), \varphi'(x_i), \dots, \varphi^{(m_i)}(x_i)$, il faut $\sum_{i=0}^{\ell} (m_i + 1) = n + 1$.

Dans ces conditions, un polynôme de degré $\leq n$ vérifiant $\varphi(x_i) = 0, \varphi'(x_i) = 0, \dots, \varphi^{(m_i)}(x_i) = 0$ devrait donc avoir un zéro d'ordre $m_i + 1$ en x_i , $i = 0, \dots, \ell$, et devrait donc admettre le facteur $\prod_{i=0}^{\ell} (x - x_i)^{m_i+1}$, impossible si $\varphi(x) \not\equiv 0$.

Cette interpolation est appelée confluente parce qu'elle résulte de la précédente quand on fait se rapprocher des points: si $x_{i+1} \rightarrow x_i$, $(\varphi(x_{i+1}) - \varphi(x_i))/(x_{i+1} - x_i)$ garde un sens et tend vers $\varphi'(x_i)$. De même, si x_{i+1} et $x_{i+2} \rightarrow x_i$, on verra que l'on peut combiner les formes relatives à x_i, x_{i+1} et x_{i+2} de façon à faire apparaître $\varphi(x_i), \varphi'(x_i)$ et $\varphi''(x_i)$. Il est clair qu'on ne fait appel qu'à des dérivées successives, à partir de l'ordre zéro.

(3) **Interpolation d'Hermite-Birkhoff.**

Ici, on se donne des dérivées quelconques en différents points, par exemple φ et φ'' en x_0 , φ' et φ^{iv} en x_1 , etc. C'est beaucoup plus difficile.

Exercice: montrez que l'on peut déterminer uniquement un élément de \mathcal{P}_4 à partir de ses dérivées premières en x_0, x_1 et x_2 distincts, et de ses valeurs en x_0 et x_1 (ce qui fait bien 5 conditions) seulement si $x_2 \neq (x_0 + x_1)/2$.

(4) **Polynômes à plusieurs variables, espaces d'éléments finis.**

On fixe des valeurs d'un polynôme et de certaines dérivées partielles en des points de \mathbb{R}^2 ou \mathbb{R}^3 . Nombreuses applications ¹. Exemple le plus simple: trouver $x_0 + x_1 x + x_1 y$

¹Cf., par exemple, J. Meinguet, Multivariate interpolation at arbitrary points made simple, *Z. Angew. Math. Phys.* **30** (1979) 292-304.

prenant des valeurs données en 3 points du plan. Condition: ces points doivent être non collinéaires.

(5) **Interpolation par moindres carrés: justification.**

Revenons à nos combinaisons S_i de fonctions $\varphi_0, \dots, \varphi_n$ approchant de mieux en mieux les valeurs d'une suite y_0, \dots, y_n au sens d'un produit scalaire discret utilisant les points x_0, \dots, x_n . S_n interpole y si le problème d'interpolation est soluble dans V_n , donc si $\det[\varphi_j(x_i)]_{i,j=0}^n \neq 0$. Or,

$$[\varphi_i(x_j)]_{i,j=0}^n \cdot [w_i \delta_{i,j}]_{i,j=0}^n \cdot \overline{[\varphi_j(x_i)]_{i,j=0}^n} = [(\varphi_i, \varphi_j)]_{i,j=0}^n,$$

matrice de Gram de la base $\{\varphi_0, \dots, \varphi_n\}$ de V_n , que l'on sait être définie positive si le produit scalaire est bien défini positif sur V_n . Voir aussi (79)

De nombreuses généralisations et extensions sont parues dans la littérature, en particulier des constructions non linéaires, comme l'interpolation par fonctions rationnelles².

Projection orthogonale et interpolation.

Soit $(f, g) = \sum_k \lambda_k(f) \overline{\lambda_k(g)}$, alors $\{L_i\} =$ base orthonormale de V .

Soit $\{b_i\}$ une base de V .

Matrice des b_i en fonction des L_k :

$$b_i = \text{interpolant} = \sum_k \lambda_k(b_i) L_k, \text{ donc } A_{i,j} = \lambda_j(b_i), \text{ et}$$

matrice de Gram $= AA^T$ définie positive $\Leftrightarrow \det A \neq 0$.

Exemples de bases de Lagrange.

Soit x_0, \dots, x_n distincts et $V_n = \mathcal{P}_n$. ℓ_j est le polynôme de degré $\leq n$ qui s'annule en $x_0, x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n$, et qui vaut 1 en $x = x_j$. On retrouve

$$\ell_j(x) = \prod_{0 \leq k \leq n, k \neq j} \frac{x - x_k}{x_j - x_k}.$$

Soit x_0, \dots, x_n distincts et examinons maintenant l'interpolation *d'Hermite* où on fixe $\varphi(x_j)$ et $\varphi'(x_j)$, $j = 0, \dots, n$: on a donc $V_n = \mathcal{P}_{2n+1}$. Les éléments de la base de Lagrange sont:

- (1) $h_j \in \mathcal{P}_{2n+1}$ tel que $h_j(x_k) = 0$ et $h_j'(x_k) = 0$, $k = 0, \dots, n$, $k \neq j$; $h_j(x_j) = 1$ et $h_j'(x_j) = 0$. Ce polynôme a donc n zéros doubles en x_k , $k = 0, \dots, n$, $k \neq j$, $h_j(x_j) = 1$, ce qui donne

$$h_j(x) = (A_j x + B_j) \prod_{0 \leq k \leq n, k \neq j} (x - x_k)^2,$$

avec

$$1 = (A_j x_j + B_j) \prod_{0 \leq k \leq n, k \neq j} (x_j - x_k)^2,$$

²Cf. J. Meinguet, On the solubility of the Cauchy interpolation problem, pp. 137-163 in A. Talbot, ed.: *Approximation Theory*, Acad. Press, 1970.

et

$$0 = A_j + 2(A_j x_j + B_j) \sum_{0 \leq k \leq n, k \neq j} \frac{1}{x_j - x_k},$$

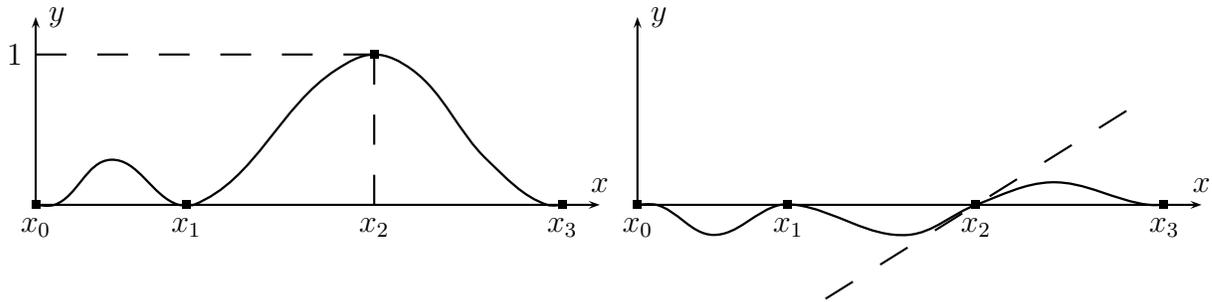
d'où

$$h_j(x) = \left[1 - 2(x - x_j) \sum_{0 \leq k \leq n, k \neq j} \frac{1}{x_j - x_k} \right] \prod_{0 \leq k \leq n, k \neq j} \left(\frac{x - x_k}{x_j - x_k} \right)^2. \quad (88)$$

- (2) $\tilde{h}_j \in \mathcal{P}_{2n+1}$ tel que $\tilde{h}_j(x_k) = 0$ et $\tilde{h}'_j(x_k) = 0$, $k = 0, \dots, n$, $k \neq j$; $\tilde{h}_j(x_j) = 0$ et $\tilde{h}'_j(x_j) = 1$. Ce polynôme a donc n zéros doubles en x_k , $k = 0, \dots, n$, $k \neq j$, et un zéro simple en x_j , avec $\tilde{h}'_j(x_j) = 1$, ce qui donne

$$\tilde{h}_j(x) = (x - x_j) \prod_{0 \leq k \leq n, k \neq j} \left(\frac{x - x_k}{x_j - x_k} \right)^2. \quad (88')$$

Exemple: graphes de h_2 et \tilde{h}_2 . Bien voir que $h'_2(x_2) = 0$:



Théorème du reste chinois. Dans le cas le plus simple, a_0, a_1, \dots, a_n sont premiers deux à deux, on reconstitue un entier p à partir des restes y_0, \dots, y_n des divisions de p par a_0, \dots, a_n , par une formule $p = \sum_{k=0}^n c_k a_0 \cdots a_{k-1} a_{k+1} \cdots a_n y_k$. C'est typiquement une formule de type Lagrange!

1.3. Interpolation polynomiale classique en formulation de Newton, différences divisées.

Si on désire construire des interpolants successifs utilisant de plus en plus de points d'un ensemble, on préfère adopter une formulation progressive:

Soit p_n l'interpolant de degré $\leq n$ en x_0, \dots, x_n . p_n doit différer de p_{n-1} d'un multiple de $(x - x_0) \cdots (x - x_{n-1})$, afin de ne pas détruire le travail d'interpolation déjà réalisé par p_{n-1} en x_0, \dots, x_{n-1} .

On a donc $p_n(x) = p_{n-1}(x) + a_{n, \text{interp}_n}(f)(x - x_0)(x - x_1) \cdots (x - x_{n-1})$, où $a_{n, \text{interp}_n}(f)$ est le coefficient de x^n de p_n , c'est d'ailleurs exactement le coefficient donné par (86), les formes de Lagrange et de Newton décrivant le même interpolant uniquement déterminé par les points d'interpolation.

$$\text{On a donc } p_n(x) = \sum_{i=0}^n a_{i, \text{interp}_i}(f)(x - x_0)(x - x_1) \cdots (x - x_{i-1}).$$

On ne calcule généralement pas les $a_{i,\text{interp}_i}(f)$ par (86), on appelle ce coefficient **différence divisée** de f en x_0, \dots, x_i , et on le note $[x_0, \dots, x_i]_f$. L'écriture (86) a cependant le mérite de montrer que $[x_0, \dots, x_n]_f$ est symétrique en x_0, \dots, x_n , mais elle ne suggère pas que cette différence divisée s'annule sur \mathcal{P}_{n-1} .

Exercice: montrez que

$$[x_i]_f = f(x_i) ; [x_i, x_j]_f = \frac{f(x_i) - f(x_j)}{x_i - x_j} ; [x_i, x_j, x_k]_f = \frac{\frac{f(x_i) - f(x_j)}{x_i - x_j} - \frac{f(x_j) - f(x_k)}{x_j - x_k}}{x_i - x_k}$$

Exercice: vérifiez $[x_i, x_j, x_k]_{x^2} = 1$.

On calcule les différences divisées par une succession de différences et de divisions (d'où leur nom) qu'on tirera de l'identité suivante: soit $p_{i,j}$ une nouvelle notation pour le polynôme interpolant f en $x_i, \dots, x_j, j \geq i$: $p_{i,j}(x_i) = f(x_i), \dots, p_{i,j}(x_j) = f(x_j)$. $p_{i,j}$ est donc de degré $\leq j - i$ (il y a $j - i + 1$ points). On a:

$$p_{i,j}(x) = \frac{(x - x_i)p_{i+1,j}(x) - (x - x_j)p_{i,j-1}(x)}{x_j - x_i}, \quad j > i. \tag{89}$$

(Identité de Neville-Aitken).

- En effet, 1. en $x = x_{i+1}, \dots, x_{j-1}, p_{i+1,j}(x) = p_{i,j-1}(x) = f(x)$, donc $p_{i,j}(x) = f(x)$;
- 2. en $x = x_i, p_{i,j-1}(x) = f(x) \Rightarrow p_{i,j}(x) = f(x)$;
- 3. en $x = x_j, p_{i+1,j}(x) = f(x) \Rightarrow p_{i,j}(x) = f(x)$.

Voyons maintenant les coefficients des plus hautes puissances de x :

$$[x_i, \dots, x_j]_f = \frac{[x_{i+1}, \dots, x_j]_f - [x_i, \dots, x_{j-1}]_f}{x_j - x_i}, \quad j > i. \tag{90}$$

Ceci permet de construire successivement les colonnes du tableau suivant, à partir de la première:

$$\begin{array}{ccccccc} [x_0]_f & & & & & & \\ [x_1]_f & [x_0, x_1]_f & & & & & \\ [x_2]_f & [x_1, x_2]_f & [x_0, x_1, x_2]_f & & & & \\ \vdots & \vdots & \vdots & \ddots & & & \\ [x_n]_f & [x_{n-1}, x_n]_f & [x_{n-2}, x_{n-1}, x_n]_f & \cdots & [x_0, \dots, x_n]_f & & \end{array}$$

On obtient un vecteur contenant tous les $[x_0, \dots, x_i]_f$ à partir des $[x_i] = f(x_i)$ par les deux boucles suivantes, la boucle intérieure exploitant le vecteur à partir de la fin:

```
for i=2:n+1;for j=n+1:-1:i;v(j)=(v(j)-v(j-1))/(x(j)-x(j-i+1));end;end;
```

Exercice.

Montrez que, si $f(x) = 1/(a - x)$, $[x_i, \dots, x_j]_f = 1/((a - x_i) \dots (a - x_j))$.

Fonctions analytiques et intégrales de contour (Hermite). Soit f analytique dans l'intérieur d'un contour C entourant x_i, \dots, x_j . Alors, $[x_i, \dots, x_j]_f$ étant une combinaison linéaire des $f(x_k) = \frac{1}{2\pi i} \int_C \frac{f(t)}{t - x_k} dt$, pour $k = i, \dots, j$,

$$[x_i, \dots, x_j]_f = \frac{1}{2\pi i} \int_C [x_i, \dots, x_j]_{(t-x)^{-1}} f(t) dt = \frac{1}{2\pi i} \int_C \frac{f(t)}{(t - x_i) \cdots (t - x_j)} dt.$$

Différences divisées des puissances de x [MilneT]. $[x_0, \dots, x_n]_{x^k}$ est évidemment nulle pour $k = 0, \dots, n-1$, et vaut 1 si $k = n$. Au-delà, appliquons la différence divisée à la fonction génératrice $\sum_0^\infty t^k x^k = \frac{1}{1-tx} = \frac{t^{-1}}{t^{-1}-x}$. D'après l'exercice plus haut,

$$[x_0, \dots, x_n]_{(1-tx)^{-1}} = \frac{t^{-1}}{(t^{-1}-x_0) \cdots (t^{-1}-x_n)} = \frac{t^n}{(1-tx_0) \cdots (1-tx_n)}.$$

La différence divisée cherchée est donc le coefficient de t^{k-n} du développement de Taylor-Maclaurin en t de $1/[(1-tx_0) \cdots (1-tx_n)]$. Il s'agit de **la fonction symétrique totale** de degré $k-n$ de $\{x_0, \dots, x_n\}$, c'est-à-dire de la somme de tous les produits (avec répétitions) de $k-n$ facteurs que l'on peut former avec x_0, \dots, x_n . $k = n+1$: $x_0 + \cdots + x_n$, $k = n+2$: $\sum_{i<j} x_i x_j + \sum x_i^2$, etc.

Lorsque les différences divisées $\mu_i = [x_0, x_1, \dots, x_{i-1}]_f$ sont connues, on calcule $p_n(x) = \sum_{i=0}^n \mu_i (x-x_0) \cdots (x-x_{i-1})$ par un algorithme comparable au schéma de Horner: `p=v(n+1); for i=n:-1:1; p=p*(xx-x(i))+v(i); end;`

Exercice. Formule de Leibniz. Montrez que

$$[x_0, \dots, x_n]_{fg} = \sum_{k=0}^n [x_0, \dots, x_k]_f [x_k, \dots, x_n]_g.$$

En effet, soient p_n et q_n les interpolants de degrés $\leq n$ de f et g en x_0, \dots, x_n . On écrit p_n et q_n selon la forme de Newton, *mais en adoptant l'ordre x_n, \dots, x_0 pour q_n* :

$$p_n(x) = \sum_{k=0}^n [x_0, \dots, x_k]_f (x-x_0) \cdots (x-x_{k-1}), \quad q_n(x) = \sum_{j=0}^n [x_j, \dots, x_n]_g (x-x_{j+1}) \cdots (x-x_n).$$

Dans le produit $p_n q_n$ terme à terme, les produits avec $j \leq k-1$ contiennent tous les facteurs de $(x-x_0) \cdots (x-x_n)$. La somme des produits avec $j \geq k$, chacun de ces produits étant de degré $n-j+k \leq n$, est donc l'interpolant de degré $\leq n$ de fg . Enfin, le coefficient de x^n est obtenu à partir de la somme des produits avec $j = k$. □

1.4. Extrapolation à la limite de Richardson.

La formule de Neville-Aitken permet de construire commodément les valeurs d'interpolants successifs en un point fixe inaccessible x à partir des valeurs de f sur une suite $x_i \rightarrow x$ quand $i \rightarrow \infty$:

on part des $c_{i,0} = f(x_i)$,

$$c_{i,k} = \frac{(x-x_i)c_{i+1,k-1}(x) - (x-x_{i+k})c_{i,k-1}}{x_{i+k} - x_i}, \quad k > 0.$$

Les $c_{i,k}$ sont les $p_{i,i+k}(x)$ de (89).

Application: $x = 0$, $x_i = 1/4^i$, $c_{i,0} = 2^i [F(u+2^{-i}) - F(u-2^{-i})]$ (formule de dérivation de Romberg).

(Justifiez le choix $x_i = 4^{-i}$).

1.5. Formulation de Newton en général.

La base de Lagrange (duale) relative aux formes $\lambda_0, \dots, \lambda_n$ est constituée de combinaisons linéaires de *emph* tous les éléments de la base $\{\varphi_0, \dots, \varphi_n\}$ donnée de V_n .

Cette fois, on réarrange les fonctions de base de façon **triangulaire** en

$$\begin{aligned} \psi_0 &= \alpha_{0,0}\varphi_0, \\ \psi_1 &= \alpha_{1,0}\varphi_0 + \alpha_{1,1}\varphi_1, \\ \dots &\dots \\ \psi_n &= \alpha_{n,0}\varphi_0 + \alpha_{n,1}\varphi_1 + \dots + \alpha_{n,n}\varphi_n, \end{aligned}$$

ainsi que les formes

$$\begin{aligned} \mu_0 &= \beta_{0,0}\lambda_0, \\ \mu_1 &= \beta_{1,0}\lambda_0 + \beta_{1,1}\lambda_1, \\ \dots &\dots \\ \mu_n &= \beta_{n,0}\lambda_0 + \beta_{n,1}\lambda_1 + \dots + \beta_{n,n}\lambda_n, \end{aligned}$$

toujours en maintenant la biorthogonalité

$$\mu_i(\psi_j) = \delta_{i,j}, \quad i, j = 0, 1, \dots, n.$$

Avantage: l'interpolant de f dans V_n est

$$p_n = \sum_{j=0}^n \mu_j(f)\psi_j,$$

puisque $\mu_i(p_n) = \sum_{j=0}^n \mu_j(f)\mu_i(\psi_j) = \mu_i(f)$, pour $i = 0, 1, \dots, n$, construction qui reste utilisable quand on passe à V_{n+1} , (**propriété de permanence**), puisque les mêmes μ_j et ψ_j se retrouvent ³ dans la solution du problème sur $V_{n+1} \times V_{n+1}^*$, il suffit d'ajouter le terme en $\mu_{n+1}\psi_{n+1}$:

$$p_{n+1} = p_n + \mu_{n+1}(f)\psi_{n+1}.$$

Ces deux réarrangements triangulaires de bases sont possibles sous les conditions suivantes:

Théorème. ⁴ *Des éléments indépendants $\varphi_0, \varphi_1, \dots$ d'un espace vectoriel X et des formes indépendantes $\lambda_0, \lambda_1, \dots$ de son dual X^* peuvent être réarrangés triangulairement*

$$\begin{array}{ll} \psi_0 = \alpha_{0,0}\varphi_0, & \mu_0 = \beta_{0,0}\lambda_0, \\ \psi_1 = \alpha_{1,0}\varphi_0 + \alpha_{1,1}\varphi_1, & \mu_1 = \beta_{1,0}\lambda_0 + \beta_{1,1}\lambda_1, \\ \psi_2 = \alpha_{2,0}\varphi_0 + \alpha_{2,1}\varphi_1 + \alpha_{2,2}\varphi_2 & \mu_2 = \beta_{2,0}\lambda_0 + \beta_{2,1}\lambda_1 + \beta_{2,2}\lambda_2, \\ \dots & \dots \end{array}$$

avec $\alpha_{i,i} \neq 0, \beta_{i,i} \neq 0, i = 0, 1, \dots$, de façon à vérifier les conditions de **biorthogonalité**

$$\mu_i(\psi_j) = \delta_{i,j}, \quad i, j = 0, 1, \dots,$$

³En toute rigueur, on aurait donc dû noter $\ell_{j,n}$ les éléments de la base de Lagrange de V_n , suggérant ainsi que la base de Lagrange de V_{n+1} n'a (normalement) aucun élément commun avec celle de V_n .

⁴[Dav] § 2.6; C. Brezinski, *Biorthogonality and its Applications to Numerical Analysis*, M. Dekker, 1992.

si les problèmes d'interpolation sur $V_n \times V_n^*$ sont tous solubles, $n = 0, 1, \dots$, c'est-à-dire si les déterminants $\det[\lambda_i(\varphi_j)]_{i,j=0}^n$ sont tous non nuls. Ce réarrangement est unique si on pose $\alpha_{0,0} = \alpha_{1,1} = \dots = 1$. On a alors

$$\psi_n = \frac{\begin{vmatrix} \lambda_0(\varphi_0) & \cdots & \lambda_0(\varphi_n) \\ \cdots & \cdots & \cdots \\ \lambda_{n-1}(\varphi_0) & \cdots & \lambda_{n-1}(\varphi_n) \\ \varphi_0 & \cdots & \varphi_n \end{vmatrix}}{\det[\lambda_i(\varphi_j)]_{i,j=0}^{n-1}}, \quad \mu_n = \frac{\begin{vmatrix} \lambda_0(\varphi_0) & \cdots & \lambda_n(\varphi_0) \\ \cdots & \cdots & \cdots \\ \lambda_0(\varphi_{n-1}) & \cdots & \lambda_n(\varphi_{n-1}) \\ \lambda_0 & \cdots & \lambda_n \end{vmatrix}}{\det[\lambda_i(\varphi_j)]_{i,j=0}^n}.$$

En effet, supposons $\psi_0, \dots, \psi_{n-1}, \mu_0, \dots, \mu_{n-1}$ déterminés, ψ_n est un élément de V_n ($n+1$ degrés de liberté) qui doit d'abord vérifier les n conditions $\mu_0(\psi_n) = \dots = \mu_{n-1}(\psi_n) = 0$. Comme μ_i est une combinaison linéaire de $\lambda_0, \dots, \lambda_i$, ces conditions sur ψ_n deviennent

$$\psi_n \in V_n : \quad \lambda_0(\psi_n) = \dots = \lambda_{n-1}(\psi_n) = 0. \quad (91)$$

Ces conditions sont bien remplies par une expression de la forme

$$\psi_n = C_n \begin{vmatrix} \lambda_0(\varphi_0) & \cdots & \lambda_0(\varphi_n) \\ \cdots & \cdots & \cdots \\ \lambda_{n-1}(\varphi_0) & \cdots & \lambda_{n-1}(\varphi_n) \\ \varphi_0 & \cdots & \varphi_n \end{vmatrix}$$

avec un scalaire quelconque C_n , puisque $\lambda_i(\psi_n)$, $i < n$, est un déterminant contenant deux lignes identiques. La dernière condition sur ψ_n est $\alpha_{n,n} = 1$, donc C_n fois cofacteur de $\varphi_n = 1 \Rightarrow C_n = 1/\det[\lambda_i(\varphi_j)]_{i,j=0}^{n-1}$.

Quant à μ_n , $\mu_n(f)$ est le coefficient de ψ_n dans le développement de l'interpolant dans V_n $p_n = \mu_n(f)\psi_n +$ éléments de V_{n-1} . Comme $\psi_n = \varphi_n +$ une combinaison linéaire d'éléments de V_{n-1} , on a donc aussi $\mu_n(f) =$ coefficient de φ_n dans l'interpolant de f dans V_n . Cet interpolant étant

$$p_n = - \frac{\begin{vmatrix} \lambda_0(\varphi_0) & \cdots & \lambda_0(\varphi_n) & \lambda_0(f) \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_n(\varphi_0) & \cdots & \lambda_n(\varphi_n) & \lambda_n(f) \\ \varphi_0 & \cdots & \varphi_n & 0 \end{vmatrix}}{\det[\lambda_i(\varphi_j)]_{i,j=0}^n},$$

on trouve bien l'expression prévue pour μ_n , et on remarque que

$$\mu_n(\varphi_0) = \dots = \mu_n(\varphi_{n-1}) = 0, \quad \mu_n(\varphi_n) = 1. \quad (92)$$

On a donc établi $\mu_i(\psi_j) = 0$ si $i < j$, $\mu_i(\psi_j) = 0$ si $i > j$, et $\mu_i(\psi_i) = 1$. □

On peut aussi faire la liaison avec la factorisation triangulaire de la matrice $[\lambda_i(\varphi_j)]_{i,j=0}^n$: $[\beta_{i,p}]_{i,p=0}^n \cdot [\lambda_p(\varphi_q)]_{p,q=0}^n \cdot [\alpha_{j,q}]_{j,q=0}^n = [\mu_i(\psi_j)]_{i,j=0}^n = [\delta_{i,j}]_{i,j=0}^n$.

Interpolation polynomiale classique en formulation de Newton, différences divisées.

Nous avons donc $V_n = \mathcal{P}_n$, $\lambda_i(f) = f(x_i)$, $i = 0, \dots, n$.

D'après (91), $\psi_n(x) = x^n + \dots$ doit s'annuler en x_0, \dots, x_{n-1} , donc

$$\psi_n(x) = (x - x_0) \dots (x - x_{n-1}).$$

D'ailleurs, l'interpolant p_n de degré $\leq n$ en x_0, \dots, x_n doit effectivement différer de p_{n-1} d'un multiple de $(x - x_0) \dots (x - x_{n-1})$, afin de ne pas détruire le travail d'interpolation déjà réalisé par p_{n-1} en x_0, \dots, x_{n-1} .

Nous savons déjà par (92) que μ_n est une combinaison linéaire de $\lambda_0, \dots, \lambda_n$ qui doit s'annuler sur \mathcal{P}_{n-1} , et vérifier $\mu_n(\psi_n) = \mu_n(x^n) = 1$. Ainsi,

$$\mu_0(f) = f(x_0); \quad \mu_1(f) = (f(x_1) - f(x_0))/(x_1 - x_0).$$

$\mu_n(f)$ étant le coefficient de x^n de l'interpolant p_n , on peut faire appel à la formulation de Lagrange

$$p_n(x) = \sum_{j=0}^n f(x_j) \ell_j(x) = \sum_{j=0}^n f(x_j) \prod_{0 \leq k \leq n, k \neq j} \frac{x - x_k}{x_j - x_k}$$

pour extraire

$$\mu_n(f) = [x_0, \dots, x_n]_f = \sum_{j=0}^n \left(f(x_j) \prod_{0 \leq k \leq n, k \neq j} \frac{1}{x_j - x_k} \right)$$

qui n'est donc autre que (86), la **différence divisée** de f en x_0, \dots, x_n .

1.6. Différences divisées et dérivées.

Formule d'Hermite-Genocchi. Si $f \in \mathcal{C}^n$, on a

$$[x_0, x_1, \dots, x_n]_f = \frac{1}{n!V_n} \int_{\Sigma_n} f^{(n)}(\lambda_0 x_0 + \lambda_1 x_1 + \dots + \lambda_n x_n) dV, \quad (93)$$

où Σ_n est le simplexe n -dimensionnel $\lambda_i \geq 0$, $\lambda_0 + \dots + \lambda_n = 1$ et $V_n = \frac{\sqrt{n+1}}{n!}$ est le volume de Σ_n (voir p. 84).

$\lambda_0, \dots, \lambda_n$ sont les **coordonnées barycentriques** de la variété Σ_n à n dimensions $\subset \mathbb{R}^{n+1}$. Une représentation paramétrique des points de Σ_n au moyen de n paramètres u_1, \dots, u_n est donnée par

$$\begin{aligned} \lambda_n &= u_n, \\ \lambda_{n-1} &= u_{n-1} - u_n, \\ \lambda_{n-2} &= u_{n-2} - u_{n-1}, \\ &\vdots \\ \lambda_1 &= u_1 - u_2, \\ \lambda_0 &= 1 - u_1. \end{aligned}$$

La formule prend alors la forme d'une intégrale n -uple

$$[x_0, x_1, \dots, x_n]_f = \int_0^1 du_1 \int_0^{u_1} du_2 \dots \int_0^{u_{n-1}} du_n f^{(n)}(x_0 + u_1(x_1 - x_0) + \dots + u_n(x_n - x_{n-1})). \quad (94)$$

THE HERMITE-GENOCCHI FORMULA by R. K. ALEXANDER. Abstract. We present an elementary proof of the Hermite-Genocchi. formula for divided differences.

<http://orion.math.iastate.edu/alex/REU/docs/HermiteGenocchi.ps>

Preuve: 1.) vrai pour $n = 1$ puisque $[x_0, x_1]_f = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{\int_{x_0}^{x_1} f'(v) dv}{x_1 - x_0} = \int_0^1 f'(x_0 + (x_1 - x_0)u) du$.

2.) Si vrai pour $n - 1$, par (90),

$$\begin{aligned} [x_0, x_1, \dots, x_n]_f &= \frac{[x_1, \dots, x_n]_f - [x_0, \dots, x_{n-1}]_f}{x_n - x_0} \\ &= \frac{1}{x_n - x_0} \int_0^1 \int_0^{u_1} \dots \int_0^{u_{n-2}} \{f^{(n-1)}(x_1 + (x_2 - x_1)u_1 + \dots + (x_n - x_{n-1})u_{n-1}) \\ &\quad - f^{(n-1)}(x_0 + (x_1 - x_0)u_1 + \dots + (x_{n-1} - x_{n-2})u_{n-1})\} du_1 \dots du_{n-1} \\ &= \frac{1}{x_n - x_0} \int_0^1 \int_0^{u_1} \dots \int_0^{u_{n-2}} \{f^{(n-1)}(x_1 + (x_2 - x_1)u_1 + \dots + (x_n - x_{n-1})u_{n-1}) \\ &\quad - f^{(n-1)}(x_1 + (x_2 - x_1)u_1 + \dots + (x_0 - x_{n-1})u_{n-1})\} du_1 \dots du_{n-1} \\ &\quad (\text{permutation } (x_0, \dots, x_{n-2}, x_{n-1}) \rightarrow (x_1, \dots, x_{n-1}, x_0)) \\ &= \frac{1}{x_n - x_0} \int_0^1 \int_0^{u_1} \dots \int_0^{u_{n-2}} \int_0^{(x_n - x_0)u_{n-1}} f^{(n)}(x_1 + (x_2 - x_1)u_1 + \dots + (x_0 - x_{n-1})u_{n-1} + v) du_1 \dots du_{n-1} dv \\ &= \int_0^1 \int_0^{u_1} \dots \int_0^{u_{n-2}} \int_0^{u_{n-1}} f^{(n)}(x_1 + (x_2 - x_1)u_1 + \dots + (x_0 - x_{n-1})u_{n-1} + (x_n - x_0)u_n) du_1 \dots du_{n-1} du_n \end{aligned}$$

avec $v = (x_n - x_0)u_n$. \square

Cas confluents. Si les x_i sont tous distincts, l'ordre dans lequel on les prend n'a aucune influence sur p_n (si on fait abstraction des erreurs d'arrondi commises dans les calculs). On peut construire un interpolant d'Hermite correct à condition de regrouper les points prenant des valeurs identiques.

Ainsi, si $x_i = x_{i+1}$, considérons la limite $x_{i+1} = x_i + h$, $h \rightarrow 0$: $[x_i]_f = f(x_i)$, $[x_{i+1}]_f = \lim_{h \rightarrow 0} f(x_i + h) = f(x_i)$, $[x_i, x_{i+1}]_f = \lim_{h \rightarrow 0} (f(x_i + h) - f(x_i))/h = f'(x_i)$, et il n'y a pas d'autre division embarrassante.

Si $x_i = \dots = x_{i+k}$, $\mu_k = [x_i, \dots, x_i]_f$ ($k+1$ fois) doit être une forme linéaire en f ne faisant intervenir que $f(x_i)$, $f'(x_i)$, \dots , $f^{(k)}(x_i)$ s'annulant sur \mathcal{P}_{k-1} et vérifiant $\mu_k(x^k) = 1$: on doit avoir

$$\underbrace{[x_i, \dots, x_i]_f}_{k+1 \text{ fois}} = f^{(k)}(x_i)/k!$$

Ceci montre qu'un interpolant tend vers une somme partielle de la série de Taylor ⁵ quand les points d'interpolation tendent les uns vers les autres.

Quand la suite $\{x_0, \dots, x_n\}$ contient plusieurs sous-suites d'éléments identiques, le tableau triangulaire des différences divisées contient des petits triangles de la forme

$$\begin{array}{ccccccc} f(x_i) & & & & & & \\ f(x_i) & f'(x_i) & & & & & \\ f(x_i) & f'(x_i) & f''(x_i)/2 & & & & \\ \vdots & \vdots & \vdots & \ddots & & & \\ f(x_i) & f'(x_i) & f''(x_i)/2 & \dots & f^{(k)}(x_i)/k! & & \end{array}$$

dont les éléments ne peuvent être calculés par différences et divisions mais doivent être introduits dans le tableau, d'ailleurs ces éléments sont des données de l'interpolation d'Hermite: quand $x_i = \dots = x_{i+k}$, les $k+1$ données relatives à ces points confluents sont bien $f(x_i), \dots, f^{(k)}(x_i)$.

Splines d'interpolation.

Soit $a = x_0 < x_1 < \dots < x_{N-1} < x_N = b$. La fonction spline d'interpolation de degré m est:

- (1) un polynôme de degré $\leq m$ dans chaque intervalle $[x_i, x_{i+1}]$, $i = 0, 1, \dots, N-1$;
- (2) interpolant en x_0, \dots, x_N ;
- (3) dans \mathcal{C}^{m-1} dans tout l'intervalle $[a, b]$.

Outre le cas primitif $m = 1$ (ligne brisée), le cas le plus rencontré est $m = 3$ (*spline cubique*). Soit $p_i \in \mathcal{P}_3$ la restriction de la fonction spline dans $[x_i, x_{i+1}]$.

On a $p(x_i) = y_i$ et $p(x_{i+1}) = y_{i+1}$ donnés;

$p'(x_i) = z_i$ et $p'(x_{i+1}) = z_{i+1}$ inconnus.

Exprimons p à partir de ces 4 spécifications:

$$\begin{aligned} p_i(x) = y_i \left[1 + \frac{2(x - x_i)}{h_i} \right] \left(\frac{x - x_{i+1}}{h_i} \right)^2 &+ y_{i+1} \left[1 - \frac{2(x - x_{i+1})}{h_i} \right] \left(\frac{x - x_i}{h_i} \right)^2 \\ &+ z_i \frac{(x - x_i)(x - x_{i+1})^2}{h_i^2} + z_{i+1} \frac{(x - x_{i+1})(x - x_i)^2}{h_i^2}, \end{aligned}$$

⁵C'est d'ailleurs comme cela que Taylor a trouvé sa série.

où $h_i = x_{i+1} - x_i$, d'après les formules (88) et (88') vues plus haut.

Exprimons maintenant la continuité de la dérivée seconde, $p_i''(x_i) = p_{i-1}''(x_i)$:

$$-6\frac{y_i}{h_i^2} + 6\frac{y_{i+1}}{h_i^2} - 4\frac{z_i}{h_i} - 2\frac{z_{i+1}}{h_i} = 6\frac{y_{i-1}}{h_{i-1}^2} - 6\frac{y_i}{h_{i-1}^2} + 2\frac{z_{i-1}}{h_{i-1}} + 4\frac{z_i}{h_{i-1}},$$

d'où

$$\frac{z_{i-1}}{h_{i-1}} + 2\left[\frac{1}{h_{i-1}} + \frac{1}{h_i}\right]z_i + \frac{z_{i+1}}{h_i} = 3\left[\frac{y_i - y_{i-1}}{h_{i-1}^2} + \frac{y_{i+1} - y_i}{h_i^2}\right],$$

$i = 1, 2, \dots, N-1$: $N-1$ équations pour les $N+1$ inconnues z_0, \dots, z_N . On fixe par exemple en plus $p_0''(a) = p_{N-1}''(b) = 0$ (fonction spline naturelle).

1.7. Reste de l'interpolation polynomiale classique.

Comme p_n interpole f en x_0, \dots, x_n , on va poser

$$f(x) - p_n(x) = k_n(x) (x - x_0) \dots (x - x_n),$$

Proposition⁶. Si $f \in \mathcal{C}^{n+1}[a, b]$, a et b réels,

$$k_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!},$$

où ξ se trouve dans le plus petit intervalle contenant x_0, \dots, x_n et x (réel).

En effet, considérons la fonction de t

$$e(t) = f(t) - p_n(t) - k_n(x) (t - x_0) \dots (t - x_n)$$

pour x fixé. La fonction e s'annule en $t = x_0, \dots, t = x_n$ et en $t = x$, ce qui fait $n+2$ points. Par le théorème de Rolle, e' s'annule au moins une fois dans chacun des $n+1$ intervalles bornés par ces $n+2$ points (et s'il y avait des points confondus, ce sont des zéros multiples de e , donc encore des zéros de e'). On poursuit jusqu'à la dérivée $(n+1)^{\text{ème}}$ de e :

$$e^{(n+1)}(t) = f^{(n+1)}(t) - p_n^{(n+1)}(t) - k_n(x) \frac{d^{n+1}}{dt^{n+1}}(t - x_0) \dots (t - x_n) = f^{(n+1)}(t) - k_n(x) (n+1)!,$$

qui doit encore s'annuler en au moins un point $t = \xi$, soit: $k_n(x) = f^{(n+1)}(\xi)/(n+1)!$. \square

La formulation de Newton donne une expression très intéressante du reste: soit encore x fixé, et considérons le polynôme q qui interpole f en x_0, \dots, x_n et x :

$$q(t) = p_n(t) + [x_0, \dots, x_n, x]_f (t - x_0) \dots (t - x_n),$$

d'où, en $t = x$,

$$f(x) - p_n(x) = [x_0, \dots, x_n, x]_f (x - x_0) \dots (x - x_n),$$

donc

$$k_n(x) = [x_0, \dots, x_n, x]_f.$$

Par (93)-(94) et le premier théorème de la moyenne, on retrouve bien

$$k_n(x) = f^{(n+1)}(\xi) \int_0^1 \int_0^{u_1} \dots \int_0^{u_{n-1}} \int_0^{u_n} du_1 du_2 \dots du_n du_{n+1} = \frac{f^{(n+1)}(\xi)}{(n+1)!}$$

Remarquons que ces formules sont parfaitement valables si x est en dehors de l'intervalle contenant x_0, \dots, x_n : l'extrapolation apparaît ici comme un cas particulier de l'interpolation.

⁶qui apparaîtra come un cas particulier du théorème de Peano.

Ref.: M.S. Floater, Error formulas for divided difference expansions and numerical differentiation, *J. Approx. Theory* **122** (2003) 1-9; C. de Boor, A divided difference expansion of a divided difference, *J. Approx. Theory* **122** (2003) 10-12.

Choix optimal des points d'interpolation.

On sait que le polynôme \hat{p}_n de \mathcal{P}_n minimisant $\|f - p\|_\infty$ sur un intervalle donné $[a, b]$ est tel que $f - \hat{p}_n$ prend alternativement les valeurs $+E_n$ et $-E_n$ en au moins $n + 2$ points de $[a, b]$ (théorème d'équioscillation), ce qui définit implicitement $n + 1$ points optimaux d'interpolation. Ces points dépendent de n et f . Un choix raisonnable de points indépendants de f ⁷ est obtenu en minimisant la norme de $\omega_{n+1}(x) = (x - x_0) \dots (x - x_n)$ sur $[a, b]$. Ceci impose

$$(\omega_{n+1}(x))_{opt} = \frac{1}{2^n} \left(\frac{b-a}{2} \right)^{n+1} T_{n+1} \left(\frac{x - (a+b)/2}{(b-a)/2} \right),$$

où on a utilisé $T_n(t) = 2^{n-1}t^n + \dots$

Exercice. Constater la divergence en norme uniforme de l'interpolation de $1/(1 + a^2x^2)$ sur des points équidistants de $[-1, 1]$ quand a est assez grand (**phénomène de Runge**).

On écrit $f(x) = (i/(2a))((x + i/a)^{-1} - (x - i/a)^{-1})$, donc $f(x) - p_n(x) = (i/(2a))([x_0, \dots, x_n, x]_{(x+i/a)^{-1}} - [x_0, \dots, x_n, x]_{(x-i/a)^{-1}})(x - x_0) \dots (x - x_n) = (i(-1)^{n+1}/(2a))((x_0 + i/a)^{-1} \dots (x_n + i/a)^{-1}(x + i/a)^{-1} - (x_0 - i/a)^{-1} \dots (x_n - i/a)^{-1}(x - i/a)^{-1})(x - x_0) \dots (x - x_n)$, d'après un exercice précédent. On examine alors l'évolution de $\omega_{n+1}(x)/\omega_{n+1}(\pm i/a) = (x - x_0) \dots (x - x_n)/((\pm i/a - x_0) \dots (\pm i/a - x_n))$ en fonction de n . Ce rapport tend rapidement vers l'infini quand $n \rightarrow \infty$ si on choisit des points équidistants sur $[-1, 1]$ et si x est près de ± 1 . Pour un traitement détaillé, voir Hairer [Hai], chap. 2.

On évite le recours à des dérivées $n^{\text{èmes}}$ dans des estimations d'erreur en évaluant la norme du projecteur d'interpolation P_{interp} : si $P_{\text{interp}}f$ est l'interpolant de f dans \mathcal{P}_n en x_0, \dots, x_n , $f - P_{\text{interp}}f = f - p - P_{\text{interp}}(f - p)$ pour $\forall p \in \mathcal{P}_n$ puisque $P_{\text{interp}}p = p$ et P_{interp} est linéaire, donc

$$|f(x) - (P_{\text{interp}}f)(x)| \leq (1 + \|P_{\text{interp}}(x)\|) \|f - \hat{p}\|,$$

où $P_{\text{interp}}(x)$ est la forme qui donne la valeur de l'interpolant en un point x fixé. En formulation de Lagrange, et en norme $\|\cdot\|_\infty$,

$$\|P_{\text{interp}}(x)\|_\infty = \max_{\|f\|_\infty \leq 1} \left| \sum_{i=0}^n f(x_i) \ell_i(x) \right| = \sum_{i=0}^n |\ell_i(x)|,$$

(**fonction de Lebesgue**). On voit donc comment l'erreur d'interpolation se compare à l'erreur de meilleure approximation.

1.8. Interpolation d'Hermite-Fejér.

L'interpolant d'Hermite-Fejér $\mathcal{H}_n f$ de $f \in \mathcal{C}[a, b]$ aux $n + 1$ points x_0, \dots, x_n est le polynôme de degré $\leq 2n + 1$ vérifiant

$$(\mathcal{H}_n f)(x_i) = f(x_i), \quad (\mathcal{H}_n f)'(x_i) = 0, \quad i = 0, \dots, n.$$

Il n'est donc pas question de faire appel à des valeurs de f' qui pourrait ne pas exister. Il s'en suit également que, si f est un polynôme non constant, $\mathcal{H}_n f$ est généralement différent de f : l'opérateur \mathcal{H}_n n'est pas un projecteur.

⁷Mais, si on sait par le théorème d'approximation de Weierstrass que $E_n \rightarrow 0$ quand $n \rightarrow \infty$, il existe toujours des fonctions continues pour lesquelles les interpolants basés sur des points donnés à l'avance ne convergent pas en norme uniforme (théorème de Bernstein-Faber).

Cependant, on peut montrer qu'avec des x_i bien choisis, les polynômes h_j de (88) sont positifs sur $[a, b]$, de sorte que $\mathcal{H}_n : (\mathcal{H}_n f)(x) = \sum_0^n f(x_j)h_j(x)$ est un opérateur positif. Comme pour les polynômes de Bernstein, on peut en tirer une démonstration du théorème de Weierstrass $\overline{\mathcal{P}} = \mathcal{C}$ selon $\| \cdot \|_\infty$ sur un compact $[a, b]$.

Exercice. Montrez que, si les x_i sont les zéros du polynôme de Tchebycheff T_{n+1} , on a

$$h_j(x) = (1 - xx_j) \left(\frac{T_{n+1}(x)}{n(x - x_j)} \right)^2, \text{ donc est positif quand } x \in [-1, 1] \text{ puisque } -1 < x_j < 1.$$

2. Formules d'intégration basées sur l'interpolation.

Une formule interpolatoire de quadrature consiste à approcher $\int_a^b f(x)d\mu(x)$ par $\int_a^b p_n(x)d\mu(x)$, où p_n est un interpolant de f . La formulation de Lagrange est la plus commode ici: la formule de quadrature est

$$\sum_{j=0}^n L_j f(x_j), \quad L_j = \int_a^b \ell_j(x)d\mu(x).$$

Les exemples les plus classiques (Simpson, etc.) sont associés à des points en progression arithmétique et seront vus au chapitre suivant.

Toute formule de ce type est une forme contenant \mathcal{P}_n dans son noyau, donc

$$\begin{aligned} \left\| \int_a^b f(x)d\mu(x) - \sum_{j=0}^n L_j f(x_j) \right\| &= \left\| \int_a^b (f(x) - \hat{p}_n(x))d\mu(x) - \sum_{j=0}^n L_j (f(x_j) - \hat{p}_n(x_j)) \right\| \\ &\leq \left[\int_a^b d\mu(x) + \sum_{j=0}^n |L_j| \right] \|f - \hat{p}_n\|_\infty, \end{aligned}$$

estimation spécialement favorable si les $L_j \geq 0$: on a alors $\sum_0^n |L_j| = \sum_0^n L_j = \int_a^b d\mu(x)$.

2.1. Reste de la formule de quadrature de Gauss.

La formule de quadrature de Gauss vue au chap. 3 p. 96 peut aussi s'interpréter comme une formule interpolatoire, mais à partir d'un interpolant d'Hermite! En effet, soit p_{2n-1} le polynôme de \mathcal{P}_{2n-1} qui vérifie $p_{2n-1}(x_j) = f(x_j)$ et $p'_{2n-1}(x_j) = f'(x_j)$, $j = 1, \dots, n$. D'après la formulation de Lagrange dans le cas confluent, on a $p_{2n-1}(x) = \sum_{j=1}^n (f(x_j)h_j(x) + f'(x_j)\tilde{h}_j(x))$ (les indices vont maintenant de 1 à n), donc $\int_a^b p_{2n-1}(x)d\mu(x) = \sum_{j=1}^n (H_j f(x_j) + \tilde{H}_j f'(x_j))$, avec $H_j = \int_a^b h_j(x)d\mu(x)$ et $\tilde{H}_j = \int_a^b \tilde{h}_j(x)d\mu(x)$. Cependant, si les x_j sont les zéros du polynôme orthogonal Φ_n de degré n relatif à $d\mu$ (et on sait [p. 93] que Φ_n a tous ses zéros distincts dans (a, b)), on a

$$\begin{aligned} \tilde{H}_j &= \int_a^b \tilde{h}_j(x) d\mu(x) \\ &= \text{constante} \int_a^b (x - x_j) \prod_{1 \leq k \leq n, k \neq j} (x - x_k)^2 d\mu(x) \\ &= \text{constante} \int_a^b \Phi_n(x) \prod_{k \neq j} (x - x_k) d\mu(x) = 0 \end{aligned}$$

par orthogonalité de Φ_n et \mathcal{P}_{n-1} !

La formule se réduit donc à une combinaison de $f(x_1), \dots, f(x_n)$ et est exacte sur \mathcal{P}_{2n-1} , ce qui suffit à l'identifier à la formule de Gauss.

Vérifions quand même que H_j est bien le $w_{j,n}$ de la p. 96:

$$\begin{aligned}
H_j &= \int_a^b h_j(x) d\mu(x) \\
&= \int_a^b [1 + A_j(x - x_j)] \prod_{1 \leq k \leq n, k \neq j} \left(\frac{x - x_k}{x_j - x_k} \right)^2 d\mu(x) && \text{(d'après (88))} \\
&= \int_a^b [1 + A_j(x - x_j)] \left(\frac{\varphi_n(x)}{(x - x_j)\varphi'_n(x_j)} \right)^2 d\mu(x) \\
&= \int_a^b \left(\frac{\varphi_n(x)}{(x - x_j)\varphi'_n(x_j)} \right)^2 d\mu(x) > 0 && \text{(orthogonalité)} \\
&= \int_a^b \left(\frac{\sum_{k=0}^{n-1} \varphi_k(x)\varphi_k(x_j)}{\beta_n \varphi'_n(x_j)\varphi_{n-1}(x_j)} \right)^2 d\mu(x) && \text{(Christoffel-Darboux)} \\
&= \frac{\sum_{k=0}^{n-1} \varphi_k^2(x_j)}{\beta_n^2 \varphi_n'^2(x_j)\varphi_{n-1}^2(x_j)} && \text{(orthonormalité)} \\
&= \frac{1}{\sum_{k=0}^{n-1} \varphi_k^2(x_j)} && \text{(Christoffel-Darboux avec } x = y = x_j)
\end{aligned}$$

Le reste de la formule est donc

$$\begin{aligned}
\int_a^b f(x) d\mu(x) - \sum_{j=1}^n H_j f(x_j) &= \int_a^b (f(x) - p_{2n-1}(x)) d\mu(x) \\
&= \int_a^b k_{2n-1}(x)(x - x_1)^2 \dots (x - x_n)^2 d\mu(x) \\
&= \int_a^b k_{2n-1}(x)\Phi_n^2(x) d\mu(x) \\
&= k_{2n-1}(\xi) \|\Phi_n\|^2 = \frac{f^{(2n)}(\xi')}{(2n)!} \|\Phi_n\|^2.
\end{aligned}$$

avec ξ et $\xi' \in [a, b]$, ayant utilisé le premier théorème de la moyenne ⁸, sachant que k_{2n-1} est continue si $n \geq 1$ (en effet, $k_{2n-1}(x) = (f(x) - p_{2n-1}(x))/\Phi_n^2(x)$ n'est susceptible d'être discontinue qu'aux zéros de Φ_n . Les limites à gauche et à droite d'un tel point $x_{j,n}$ sont, comme $p = f$ et $p' = f'$ en ce point, $(f''(x_{j,n}) - p''_{2n-1}(x_{j,n}))/ (2\Phi_n'^2(x_{j,n}))$).

2.2. Formules de quadratures de Gauss avec points imposés.

Fixons $x_0^{(1)}, \dots, x_{n_1}^{(1)}$, laissons $x_1^{(2)}, \dots, x_{n_2}^{(2)}$ libres de façon à avoir une formule $\int_a^b f(x) d\mu(x) \approx \sum_0^{n_1} h_i^{(1)} f(x_i^{(1)}) + \sum_1^{n_2} h_j^{(2)} f(x_j^{(2)})$ exacte pour des polynômes de degré aussi élevé que possible (*degré de précision*). Comme il y a $n_1 + 2n_2 + 1$ degrés de liberté (les $n_1 + 1$ $h_i^{(1)}$, les n_2 $x_j^{(2)}$ et les n_2 $h_j^{(2)}$), on s'attend à un degré de précision $n_1 + 2n_2$. Solution: soit $\omega_1(x) = (x - x_0^{(1)}) \dots (x - x_{n_1}^{(1)})$. On prend $x_j^{(2)}$ et $h_j^{(2)}\omega_1(x_j^{(2)}) =$ noeuds et poids de la formule

⁸J. Mawhin, *Introduction à l'analyse*, Cabay, 1979, p.334; *Analyse. Fondements, techniques, évolution*, De Boeck, p. 388 dans la 2^{ème} édition de 1997: si f est réelle continue sur $[a, b]$, g positive sur $[a, b]$, g et fg intégrables sur $]a, b]$, alors $\exists c \in [a, b] : \int_a^b fg = f(c) \int_a^b g$

de Gauss relative à la mesure $\omega_1(x) d\mu(x)$. Alors, $\forall p \in \mathcal{P}_{n_1+2n_2}$, $p = q\omega_1 + r$, avec $q \in \mathcal{P}_{2n_2-1}$ et $r \in \mathcal{P}_{n_1}$. On a bien

$$\int_a^b p(x) d\mu(x) = \int_a^b q(x)\omega_1(x) d\mu(x) + \int_a^b r(x) d\mu(x) = \underbrace{\sum_1^{n_2} h_j^{(2)} \omega_1(x_j^{(2)}) q(x_j^{(2)})}_{\text{exacte par Gauss}} + \sum_0^{n_1} h_i^{(1)} r(x_i^{(1)}) + \sum_1^{n_2} h_j^{(2)} r(x_j^{(2)}),$$

où il suffit donc de fixer les $n_1 + 1$ $h_i^{(1)}$ pour que la formule soit exacte pour tout $r \in \mathcal{P}_{n_1}$.

On distingue les règles de

- (1) Lobatto: $n_1 = 1$, $x_i^{(1)} = a$ et b .
- (2) Radau: $n_1 = 0$, $x_0^{(1)} = a$ ou b .
- (3) Kronrod: $n_1 = n_2 = n$, les $x_i^{(1)}$ proviennent d'une formule de Gauss. La théorie est assez délicate, la mesure $\omega_1(x) d\mu(x)$ n'étant plus de signe constant (théorie des polynômes de Stieltjes).

Cf. W. Gautschi, Orthogonal polynomials and quadrature, *ETNA* (<http://etna.mcs.kent.edu/html/>) **9** (1999) 65-76.

On se sert souvent d'une formule de Kronrod pour estimer l'exactitude du résultat obtenu avec la règle de Gauss sous-jacente. Si l'écart est supérieur à une tolérance demandée, on reprend avec un intervalle plus petit, voir règles adaptatives plus bas.

2.3. Points de Tchebycheff: règle de Clenshaw-Curtis.

Si on dispose de l'approximation discrète $\sum_0^N c_k^{(N)}(f) T_k$ de f , en fait interpolation de f aux $N + 1$ extrema $\cos(j\pi/N)$, $j = 0, \dots, N$ de T_N , on estime $\int_{-1}^1 f(x) dx$ par

$$\sum_0^N c_k^{(N)}(f) \int_{-1}^1 T_k(x) dx = \sum_{0 \leq k \text{ pair} \leq N} c_k^{(N)}(f) \frac{2}{1-k^2}, \text{ ce qui donne, compte tenu de}$$

$$c_k^{(N)}(f) = \frac{2}{N} \sum_0^N \cos \frac{jk\pi}{N} f \left(\cos \frac{j\pi}{N} \right),$$

$$\int_{-1}^1 f(x) dx \approx \sum_{j=0}^N \gamma_j^{(N)} f \left(\cos \frac{j\pi}{N} \right), \quad \gamma_j^{(N)} = \frac{2}{N} \sum_{0 \leq k \text{ pair} \leq N} \frac{2 \cos(jk\pi/N)}{1-k^2}.$$

cf. Is Gauss quadrature better than Clenshaw-Curtis? (SIAM Annual Meeting, July 2006) <http://web.comlab.ox.ac.uk/>,
Is Gauss quadrature better than Clenshaw-Curtis? (SIAM Review, to appear) <http://web.comlab.ox.ac.uk/oucl/work/>

2.4. Règles adaptatives d'intégration.

On ne subdivise un intervalle que si une estimation de l'erreur dépasse une tolérance imposée. On arrive à une construction récursive du type

```
integ(f,a,b,tol)
  f1=formul1(f,a,b) /* formule de base sur (a, b) */
  f2=formul2(f,a,b) /* formule plus raffinée, par exemple, Kronrod */
  if |f2-f1| < = tol then return f1
  else return integ(f,a,(a+b)/2,tol/2) + integ(f,(a+b)/2,b,tol/2)
```

end Cf. J.R. Rice, *Numerical Methods, Software, and Analysis*, McGraw-Hill, 1983, pp. 193-198; P. Favati, G. Lotti, F. Romani, Algorithm 691; Improving QUADPACK automatic integration routines, *ACM Trans. Math. Soft.* **17** (1991) 218-232.

Pour des développements récents, en particulier sur l'intégration numérique de fonctions de plusieurs variables, voir le très dynamique groupe "NINES" de la KULeuven,

<http://www.cs.kuleuven.ac.be/cwis/research/nines/>

3. Représentation du reste: théorème de Peano.

Lorsqu’une formule approchée est exacte pour des polynômes de degré n , on essaye le plus souvent de trouver pour le reste une expression de la forme $Cf^{(n+1)}(\xi)$, le prototype de ce genre de formule étant le reste d’une approximation de Taylor (en un point fixé x):

$$R_{\text{Taylor},n}f := f(x) - f(x_0) - (x - x_0)f'(x_0) - \dots - \frac{(x - x_0)^n f^{(n)}(x_0)}{n!} = \frac{(x - x_0)^{n+1} f^{(n+1)}(\xi)}{(n + 1)!},$$

($\xi \in [x_0, x]$), valable si $f \in \mathcal{C}^{n+1}[x_0, x]$. On a trouvé une extension au reste de l’interpolation polynomiale:

$$R_{\text{interp},n}f := f(x) - \text{interpolant de } f \text{ en } x_0, \dots, x_n = \frac{(x - x_0) \dots (x - x_n) f^{(n+1)}(\xi)}{(n + 1)!},$$

avec $\xi \in [\min(x_0, \dots, x_n, x), \max(x_0, \dots, x_n, x)]$.

Deux problèmes:

- (1) L’incertitude sur ξ peut conduire à surestimer des bornes d’erreur. Cette situation ne peut que s’aggraver si on utilise de telles estimations dans d’autres formules, par exemple dans une règle de quadrature interpolatoire.
- (2) On peut vouloir appliquer la formule à $f \notin \mathcal{C}^{n+1}$.

Et deux idées pour s’en sortir:

- (1) Comme on s’intéresse à des formes linéaires, $Rf = R(f - p)$ pour tout p de degré $\leq n$, par exemple, une meilleure approximation, mais aussi toute autre approximation, et pas nécessairement de degré n , une approximation de Taylor, etc.
- (2) On connaît aussi une formulation intégrale du reste de l’approximation de Taylor de n’importe quel degré $m \leq n$

$$R_{\text{Taylor},m}f = \int_{x_0}^x \frac{(x - t)^m f^{(m+1)}(t)}{m!} dt.$$

Le théorème de Peano incorpore ces deux idées dans l’énoncé remarquable suivant:

3.1. Théorème (Peano). *Soit R une forme sur $\mathcal{C}^r(a, b)$ s’annulant sur \mathcal{P}_m . Alors, si $f \in \mathcal{C}^{m+1}(a, b)$, ($m \geq r \geq 0$),*

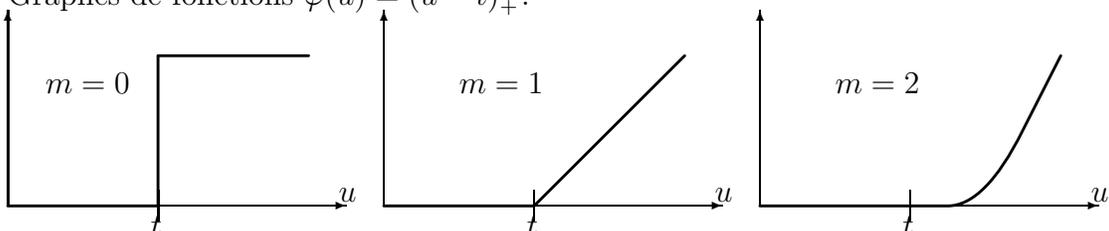
$$Rf = \int_a^b K_m(t) f^{(m+1)}(t) dt, \tag{95}$$

avec $K_m(t) = R \frac{(\cdot - t)_+^m}{m!}$, c’est-à-dire pour t fixé, $K_m(t) = R$ appliquée à la fonction de u qui vaut 0 tant que $u < t$, et qui vaut $(u - t)^m / m!$ quand $u \geq t$.

De plus, si K_m est de signe constant sur (a, b) ,

$$Rf = C f^{(m+1)}(\xi), \quad \text{où } C = R \frac{u^{m+1}}{(m + 1)!}.$$

Graphes de fonctions $\varphi(u) = (u - t)_+^m$:



En effet, R s'applique à $\mathcal{C}^r \Rightarrow R$ est une combinaison linéaire de valeurs ponctuelles de $f, f', \dots, f^{(r)}$ et d'intégrales de $f, f', \dots, f^{(r)}$:

$$Rf = \sum_i \sum_{j \leq r} c_{i,j} f^{(j)}(x_{i,j}) + \sum_{j \leq r} \int_a^b d_j(u) f^{(j)}(u) du$$

(notons que les valeurs ponctuelles pourraient être incorporées dans des intégrales au sens de Riemann-Stieltjes).

on a donc $Rf = R(f - p)$ avec n'importe quel $p \in \mathcal{P}_m$, choisissons $p =$ approximation de Taylor de degré m autour de a , donc $f(x) - p(x) = \int_a^x \frac{(x-t)^m f^{(m+1)}(t)}{m!} dt$, ce qui s'écrit encore comme l'intégrale sur (a, b) : $\int_a^b \frac{(x-t)_+^m f^{(m+1)}(t)}{m!} dt$,

remarquons aussi que la $j^{\text{ème}}$ dérivée de $f - p$ est $f^{(j)}(x) - p^{(j)}(x) = \int_a^b \frac{(x-t)_+^{m-j} f^{(m+1)}(t)}{(m-j)!} dt$, donc

$$\begin{aligned} Rf = R(f - p) &= \sum_i \sum_{j \leq r} c_{i,j} \int_a^b \frac{(x_{i,j} - t)_+^{m-j} f^{(m+1)}(t)}{(m-j)!} dt + \\ &\quad \sum_{j \leq r} \int_a^b d_j(u) \int_a^b \frac{(u-t)_+^{m-j} f^{(m+1)}(t)}{(m-j)!} dt du \\ &= \int_a^b \left\{ \sum_i \sum_{j \leq r} c_{i,j} \frac{(x_{i,j} - t)_+^{m-j}}{(m-j)!} + \sum_{j \leq r} \int_a^b d_j(u) \frac{(u-t)_+^{m-j}}{(m-j)!} du \right\} f^{(m+1)}(t) dt \end{aligned}$$

(théorème de Fubini-Tonelli pour les fonctions intégrables), ce qui est bien $\int_a^b K_m(t) f^{(m+1)}(t) dt$

avec $K_m(t) = \sum_i \sum_{j \leq r} c_{i,j} \frac{(x_{i,j} - t)_+^{m-j}}{(m-j)!} + \sum_{j \leq r} \int_a^b d_j(u) \frac{(u-t)_+^{m-j}}{(m-j)!} du = R$ appliqué à $\varphi(u) = (u-t)_+^m / (m!)$ pour t fixé $\in [a, b]$.

Enfin, si K_m est de signe constant sur $[a, b]$, $\int_a^b K_m(t) f^{(m+1)}(t) dt = C f^{(m+1)}(\xi)$, avec $C = \int_a^b K_m(t) dt$, par le théorème de la moyenne du calcul intégral (voir note p. 160). Par conséquent, si on applique R à un polynôme f tel que $f^{(m+1)}(x) \equiv 1$, comme $f(x) = x^{m+1} / (m+1)!$, on a bien $Rf = C$. \square

Remarques.

- (1) Pour t fixé dans $[a, b]$ et $m \geq 1$, $\varphi_m(u) = (u-t)_+^m / m!$ est la $m^{\text{ème}}$ primitive de la fonction échelon $\varphi_0(u) = (1 + \text{sign}(u-t))/2$, avec $\varphi_m(a) = 0$: $\varphi_m(u) = \int_a^u \varphi_{m-1}(t) dt$.
- (2) Au sens des distributions, φ_m est la $(m+1)^{\text{ème}}$ primitive de la distribution de Dirac $\delta(u-t)$.
- (3) Si on admet la forme (95), on retrouve effectivement une valeur de K_m , disons $K_m(t_0)$, en appliquant R à une fonction dont la dérivée $(m+1)^{\text{ème}}$ est une distribution de Dirac, précisément $f(t) = (t-t_0)_+^m / m!$.
- (4) Si $m > r$, $K_m(a) = K_m(b) = 0$: $\varphi_m \in \mathcal{C}^{m-1}$, donc K_m est continue (ne contient que des dérivées au plus $r^{\text{èmes}}$ de φ_m), $K_m(a) = R$ appliqué au polynôme $(u-a)^m / m! \in \mathcal{P}_m$, $K_m(b) = R$ appliqué à la fonction nulle.

- (5) Si $m > r$, $K_m(t) = -\int_a^t K_{m-1}(u) du$, puisque $K_m(a) = K_m(b) = 0$ et $Rf = \int_a^b K_m(t) f^{(m+1)}(t) dt \Rightarrow Rf = [K_m f^{(m)}]_a^b - \int_a^b K'_m(t) f^{(m)}(t) dt$.
- (6) Si $Rf = 0$ pour $f \in \mathcal{P}_n$, K_0, \dots, K_{n-1} doivent changer de signe sur $[a, b]$ puisque $\int_a^b K_m(t) dt = 0$ tant que $m < n$ (il suffit de prendre $f(t) = t^{m+1}$ dans (95)). Seul K_n a une chance de ne pas changer de signe sur $[a, b]$ si Rf ne s'annule pas sur \mathcal{P}_{n+1} .

Exemples:

Différence divisée: d'après (93), $K_n(t)$ = la mesure de l'intersection du simplexe S_n et de l'hyperplan $\lambda_0 x_0 + \dots + \lambda_n x_n = t$. K_n est donc une fonction positive, d'intégrale définie $1/n!$. On a

$$K_n(t) = B_n(t; x_0, \dots, x_n),$$

fonction de classe \mathcal{C}^{n-2} , dont la restriction entre deux x_i est un polynôme de \mathcal{P}_{n-1} (**B-spline**.) En effet, 1.) pour $n = 1$, on a $[x_0, x_1]_f = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \int_{-\infty}^{\infty} \frac{\chi((x_0, x_1))}{x_1 - x_0} f'(t) dt$,

donc B_1 est une fonction discontinue, constante par intervalles, d'intégrale définie unité; 2.)

passage de $n - 1$ à n : $[x_0, \dots, x_n]_f = \frac{[x_1, \dots, x_n]_f - [x_0, \dots, x_{n-1}]_f}{x_n - x_0} = \int_{-\infty}^{\infty} \frac{B_{n-1}(t; x_1, \dots, x_{n-1}) - B_{n-1}(t; x_0, \dots, x_{n-1})}{x_n - x_0} f^{(n-1)}(t) dt =$

$\int_{-\infty}^{\infty} B_n(t; x_0, \dots, x_n) f^{(n)}(t) dt$. B_n est donc une primitive d'une combinaison de fonctions B_{n-1} , donc d'un ordre de continuité plus élevé, de degré plus élevé que B_{n-1} entre deux x_i .

Formules d'intégration: voir p. 171

CHAPITRE 5

Différences finies.

Interpolation, dérivation, intégration, formules sommatoires.

On s'occupe ici de traiter des fonctions à partir de valeurs en des points en progression arithmétique $x_0, x_0 \pm h, x_0 \pm 2h, \dots$

Les manipulations de différences de valeurs de fonctions ont d'abord servi à l'utilisation de tables, un art aujourd'hui presque éteint.

29 DEGRÉS						Parties prop.	
'	Sin.	D	Tang.	D.C	Cotg.	"	23
...	10	3,8
10	1,68 784	22	1,74 673	30	0,25 327	20	7,7
11	807	23	702	29		30	11,5
12	829	22	732	30		40	15,3
	etc.					50	19,2
						6	2,3
						7	2,7
						8	3,1
						9	3,5

Extrait des *Tables de logarithmes à cinq décimales et autres tables*, par N.J. Schons, La Procure-Casterman, 4^{ème} éd., 1959.

Des logarithmes décimaux de sinus et tangentes sont donnés de minute en minute, les nombres négatifs sont donnés sous la forme $\bar{x}, yyyyy$ signifiant $-x + 0.yyyyy$. Les différences premières (“D.C” signifie que ces différences valent également pour les cotangentes) suffisent à reconstituer une valeur pour un angle jusqu’à la précision de la seconde d’arc. Les colonnes “parties proportionnelles” servent à faciliter les calculs d’interpolation linéaire.

Le calcul aux différences finies s’est rapidement développé en même temps que le calcul infinitésimal (dans la dénomination, “finies” s’oppose à “infinitésimales”), ce qui est l’occasion de rencontrer des correspondances intéressantes (il y a des formules de sommation par parties, etc.) On retrouve ce calcul dans certains domaines de l’algèbre (groupes particuliers), de la théorie des fonctions (transformations de fonctions) et, bien entendu, en mathématiques discrètes.

Nombreuses applications en physique et en théorie du signal.

1. Les opérateurs du calcul aux différences.

Pour h fixé, on envisage d’appliquer les opérateurs suivants à des fonctions définies sur un ensemble contenant au moins des nombres en progression arithmétique de pas h :

- (1) **Opérateur de translation.**
 $E : (Ef)(x) = f(x + h).$
- (2) **Opérateur de différence progressive.**
 $\Delta : (\Delta f)(x) = f(x + h) - f(x).$
- (3) **Opérateur de différence régressive.**
 $\nabla : (\nabla f)(x) = f(x) - f(x - h).$
- (4) **Opérateur de différence centrale.**
 $\delta : (\delta f)(x) = f(x + h/2) - f(x - h/2).$
- (5) **Opérateur de moyenne.**
 $\mu : (\mu f)(x) = \frac{f(x + h/2) + f(x - h/2)}{2}.$
- (6) **Opérateur de dérivation.** $D : (Df)(x) = f'(x).$
- (7) **Opérateur d'intégration.** $J : (Jf)(x) = \int_x^{x+h} f(t)dt.$

Les deux derniers opérateurs sont bien sûr des opérateurs de l'analyse infinitésimale. On verra comment les approcher au moyen des opérateurs aux différences proprement dits.

Tous ces opérateurs sont bien des applications linéaires définies sur un ensemble très large de fonctions. On peut donc considérer des sommes et produits de ces opérateurs, ainsi:

$$\begin{aligned} (E\Delta f)(x) &= f(x + 2h) - f(x + h), \\ (\Delta^2 f)(x) &= f(x + 2h) - 2f(x + h) + f(x), \\ (\delta^2 f)(x) &= (\Delta\nabla f)(x) = (\nabla\Delta f)(x) = f(x + h) - 2f(x) + f(x - h). \end{aligned}$$

On vérifie que le *produit est commutatif*.

Accessibilité. Si on ne dispose que des valeurs $f(x_0), f(x_0 \pm h), f(x_0 \pm 2h), \dots$, toute formule utilisable devra nécessairement aboutir à une combinaison de ces valeurs disponibles.

Ainsi, on peut librement disposer de n'importe quelle puissance (entière) et produits de E, Δ et ∇ en toute abscisse $x_0 + kh$, avec k entier.

Par contre, δ et μ doivent être appliqués à des abscisses $x_0 + (k + 1/2)h$, k entier. δ^2 et μ^2 de f en x font appel à $f(x)$ et $f(x \pm h)$ et peuvent donc s'appliquer à nouveau à des abscisses $x_0 + kh$, k entier. Il en est de même pour tout produit $\mu^i \delta^j$ si $i + j$ est pair.

Les mêmes différences calculables s’expriment indifféremment en termes de différences progressives, centrales ou régressives, soit $x_k = x_0 + kh$:

$$\begin{array}{r}
 f(x_{-3}) \\
 f(x_{-2}) \\
 f(x_{-1}) \\
 f(x_0) \\
 f(x_1) \\
 f(x_2) \\
 f(x_3)
 \end{array}
 \begin{array}{l}
 (\Delta f)(x_{-2}) = (\delta f)(x_{-3/2}) = (\nabla f)(x_{-1}) \\
 (\Delta f)(x_{-1}) = (\delta f)(x_{-1/2}) = (\nabla f)(x_0) \\
 (\Delta f)(x_0) = (\delta f)(x_{1/2}) = (\nabla f)(x_1) \\
 (\Delta f)(x_1) = (\delta f)(x_{3/2}) = (\nabla f)(x_2) \\
 \\
 (\Delta^2 f)(x_{-2}) = (\delta^2 f)(x_{-1}) = (\nabla^2 f)(x_0) \\
 (\Delta^2 f)(x_{-1}) = (\delta^2 f)(x_0) = (\nabla^2 f)(x_1) \\
 (\Delta^2 f)(x_0) = (\delta^2 f)(x_1) = (\nabla^2 f)(x_2) \\
 (\Delta^2 f)(x_1) = (\delta^2 f)(x_2) = (\nabla^2 f)(x_3) \\
 \\
 (\Delta^3 f)(x_{-2}) = (\delta^3 f)(x_{-1/2}) = (\nabla^3 f)(x_1) \\
 (\Delta^3 f)(x_{-1}) = (\delta^3 f)(x_{1/2}) = (\nabla^3 f)(x_2) \\
 (\Delta^3 f)(x_0) = (\delta^3 f)(x_1) = (\nabla^3 f)(x_2) \\
 (\Delta^3 f)(x_1) = (\delta^3 f)(x_2) = (\nabla^3 f)(x_3)
 \end{array}
 \tag{96}$$

Inversion et autres identités.

E^k signifie clairement $(E^k f)(x) = f(x + kh)$ pour tout $k \in \mathbb{Z}$. Tout naturellement, on définira $(E^s f)(x) = f(x + sh)$ pour s quelconque. On peut alors exprimer tous les opérateurs aux différences en fonction de l’un d’entre eux, soit E :

$$\Delta = E - 1, \quad \nabla = 1 - E^{-1}, \quad \delta = E^{1/2} - E^{-1/2}, \quad \mu = \frac{E^{1/2} + E^{-1/2}}{2}, \quad E^{\pm 1} = (\mu \pm \delta/2)^2, \quad 1 = \mu^2 - \delta^2/4.$$

(pour D et J , on verra un peu plus loin).

L’inverse de Δ n’est définie qu’à une constante près, comme l’inverse de D : ce doit donc être une sorte de primitive discrète, on y reviendra.

Remarquons que ΔD^{-1} est bien défini: ce n’est autre que J !

Ce calcul opérationnel est partiellement justifié par l’effet de l’application de ces opérateurs aux fonctions exponentielles $e^{\lambda x}$: on constate que l’on retrouve la même exponentielle multipliée par une constante, que les exponentielles sont donc des *fonctions propres* de ces opérateurs, avec les *valeurs propres*

$$\begin{array}{ccccccc}
 E & \Delta & \nabla & \delta & \mu & D & J \\
 e^\theta & e^\theta - 1 & 1 - e^{-\theta} & e^{\theta/2} - e^{-\theta/2} & \frac{e^{\theta/2} + e^{-\theta/2}}{2} & \frac{\theta}{h} & h \frac{e^\theta - 1}{\theta}
 \end{array}
 \tag{97}$$

$$= 2 \sinh(\theta/2) \quad = \cosh(\theta/2)$$

où $\theta = h\lambda$.

Autres opérateurs.

q -différence: $(D_q f)(x) = \frac{f(qx) - f(x)}{(q - 1)x}$.

Opérateur d’Askey-Wilson: choisissons le plus général de fonctions φ_1 et φ_2 telles que $(\mathcal{D}_{AW} f)(x) = \frac{f(\varphi_2(x)) - f(\varphi_1(x))}{\varphi_2(x) - \varphi_1(x)}$ envoie \mathcal{P}_n dans \mathcal{P}_{n-1} .

Cf. R. Askey, J. Wilson: Some basic hypergeometric orthogonal polynomials that generalize Jacobi polynomials, *Memoirs of the AMS* **54**, n° 319 (1985). G. Gasper, M. Rahman: *Basic Hypergeometric Series*, Encyc. of Math. and Applic. **35**, Cambridge U.P., 1990. R. Koekoek & R.F. Swarttouw : The Askey-scheme

of hypergeometric orthogonal polynomials and its q-analogue. Report Techn. Univ. Delft no. 98-17, 1998, voir p. 108.

2. Interpolation.

Estimer $f(x_0 + sh)$ à partir des $f(x_0 + kh)$, $k \in \mathbb{Z}$ revient donc à exprimer E^s à partir de puissances entières des opérateurs disponibles, en fait à partir des expressions du tableau (96)
Séries de puissances de Δ .

Développons $E^s = (1 + \Delta)^s$ selon le binôme de Newton généralisé:

$$E^s = (1 + \Delta)^s = \sum_{k=0}^{\infty} \binom{s}{k} \Delta^k = 1 + s\Delta + \frac{s(s-1)}{2} \Delta^2 + \frac{s(s-1)(s-2)}{6} \Delta^3 + \dots$$

(**formule de Newton-Gregory**).

Deux façons de justifier ce développement:

- (1) Toute somme partielle représente un opérateur d'interpolation polynomiale, selon la forme de Newton, en effet, avec $x = x_0 + sh$ et $x_i = x_0 + ih$, $i = 0, 1, \dots, s(s-1) \dots (s-k+1) = (x-x_0)(x-x_1) \dots (x-x_{k-1})/h^k$ et, en comparant la formation des différences finies aux différences divisées:

$$(\Delta^k f)(x_p) = (\nabla^k f)(x_{p+k}) = (\delta^k f)(x_{p+k/2}) = k!h^k [x_p, x_{p+1}, \dots, x_{p+k}]_f = \sum_{j=0}^k (-1)^j \binom{k}{j} f(x_j). \tag{98}$$

En effet, 1) vrai pour $k = 0$; 2) si vrai en $k - 1$,

$$\begin{aligned} (\Delta^k f)(x_p) &= (\Delta^{k-1} f)(x_{p+1}) - (\Delta^{k-1} f)(x_p) \\ &= (k-1)!h^{k-1} \underbrace{([x_{p+1}, x_{p+2}, \dots, x_{p+k}]_f - [x_p, x_{p+1}, \dots, x_{p+k-1}]_f)}_{(x_{p+k}-x_p=kh)[x_p, x_{p+1}, \dots, x_{p+k}]_f} \end{aligned}$$

La formule de Newton-Gregory est donc *exacte* lorsqu'elle s'applique à un polynôme, cas où la série se limite à un nombre fini de termes non nuls.

- (2) Lorsqu'on applique la formule à une exponentielle $e^{\lambda x}$, on voit qu'on est en présence de la série numérique $e^{s\theta} = [1 + (e^\theta - 1)]^s = \sum_0^\infty \binom{s}{k} (e^\theta - 1)^k$, avec $\theta = h\lambda$, série qui converge si $|e^\theta - 1| < 1$. On peut étudier toute fonction susceptible d'une représentation du type $f(x) = \int_L e^{\lambda x} \varphi(\lambda) d\lambda$, forme que l'on trouve dans des formules d'inversion de transformées de Laplace et de Fourier.

Exercice: constater la divergence de la série de Newton-Gregory de e^s à partir des différences de la suite $\{1, e, e^2, e^3, \dots\}$.

Il y a également une forme régressive développant $E^s = (1 - \nabla)^{-s}$.

Séries de puissances de δ .

Les sommes partielles successives de la série de Newton-Gregory sont des valeurs en $x = x_0 + sh$ d'interpolants en x_0 ; x_0 et x_1 ; x_0, x_1 et x_2 , etc. Supposons s entre 0 et 1: il faudrait alors plutôt partir d'un certain x_{-p} , et commencer à rencontrer des valeurs significatives à partir de la $p^{\text{ème}}$ somme partielle.

Il vaut mieux prendre les points d'interpolation dans l'ordre $x_0, x_1, x_{-1}, x_2, \dots$, ce qui donne

$[x_0]_f + [x_0, x_1]_f(x - x_0) + [x_0, x_1, x_{-1}]_f(x - x_0)(x - x_1) + [x_0, x_1, x_{-1}, x_2]_f(x - x_0)(x - x_1)(x - x_{-1}) + \dots$,
 donc, par (98),

$$E^s = 1 + s\delta E^{1/2} + \frac{s(s-1)}{2}\delta^2 + \frac{s(s-1)(s+1)}{6}\delta^3 E^{1/2} + \frac{s(s-1)(s+1)(s-2)}{24}\delta^4 + \dots$$

(formule de Gauss progressive).

Si on prend plutôt l'ordre $x_0, x_{-1}, x_1, x_{-2}, x_2, \dots$, on a

$$E^s = 1 + s\delta E^{-1/2} + \frac{s(s+1)}{2}\delta^2 + \frac{s(s+1)(s-1)}{6}\delta^3 E^{-1/2} + \frac{s(s+1)(s-1)(s+2)}{24}\delta^4 + \dots$$

(formule de Gauss régressive).

Gauss interpolation formula <http://eom.springer.de/g/g043440.htm>

Steffensen interpolation formula <http://eom.springer.de/s/s087620.htm>

Formule de Stirling: moyenne des deux formules précédentes

$$E^s = 1 + s\mu\delta + \frac{s^2}{2}\delta^2 + \frac{s(s^2-1)}{6}\mu\delta^3 + \frac{s^2(s^2-1)}{24}\delta^4 + \dots$$

Formule de Bessel: moyenne de la formule de Gauss progressive et de E fois la formule régressive pour E^{s-1} :

$$E^s = \frac{1+E}{2} + (s-1/2)\delta E^{1/2} + \frac{s(s-1)}{2}\delta^2 \frac{1+E}{2} + \frac{s(s-1)(s-1/2)}{6}\delta^3 E^{1/2} + \frac{s(s-1)(s+1)(s-2)}{24}\delta^4 \frac{1+E}{2}$$

Formule d'Everett: formule précédente où on remplace les $\delta^{2p+1}E^{1/2}$ par $\delta^{2p}(E-1)$:

$$\begin{aligned} E^s &= \frac{1+E}{2} + (2s-1)\frac{E-1}{2} + \frac{s(s-1)}{2}\delta^2 \frac{1+E}{2} + \frac{s(s-1)(2s-1)}{6}\delta^2 \frac{E-1}{2} + \frac{s(s-1)(s+1)(s-2)}{24}\delta^4 \\ &= 1 - s + sE - \frac{s(s-1)(s-2)}{6}\delta^2 + \frac{(s+1)s(s-1)}{6}\delta^2 E + \dots \\ &= \sum_{k=0}^{\infty} \left[-\binom{s+k-1}{2k+1} \delta^{2k} + \binom{s+k}{2k+1} \delta^{2k} E \right]. \end{aligned}$$

3. Dérivation.

Il s'agit d'extraire D en fonction de E , ou Δ , etc.

Ecrivons symboliquement la formule de Taylor:

$$(Ef)(x) = f(x+h) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x) h^k}{k!} = \sum_{k=0}^{\infty} \frac{h^k}{k!} (D^k f)(x) = (e^{hD} f)(x),$$

soit,

$$E = e^{hD}, \quad D = \frac{1}{h} \ln E.$$

Formules progressives et régressives (**Markoff**):

$$hD = \ln(1+\Delta) = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{\Delta^k}{k} = -\ln(1-\nabla) = \sum_{k=1}^{\infty} \frac{\nabla^k}{k}.$$

Formules centrées (**Bickley**): d'après (97), $hD = 2 \arg \sinh(\delta/2) = 2 \int_0^{\delta/2} (1+u^2)^{-1/2} du = 2 \sum_{k=0}^{\infty} \frac{(-1/2)(-3/2)\dots(1/2-k)}{k!} \frac{(\delta/2)^{2k+1}}{2k+1}$,

$$(hD)^n = \delta^n \left[1 - \frac{n}{24} \delta^2 + \frac{5n^2 + 22n}{5760} \delta^4 + \dots \right],$$

utilisable en x_0 seulement si n est pair. Si n est impair, on s'arrange pour avoir des termes en $\mu\delta^n$, $\mu\delta^{n+2}$, par $(1 + \delta^2/4)^{1/2} = \mu$:

$$(hD)^n = \mu\delta^n \left[1 - \frac{n+3}{24} \delta^2 + \frac{5n^2 + 52n + 175}{5760} \delta^4 + \dots \right].$$

Pour évaluer $f^{(n)}(x_0 + sh)$: multiplier D^n par un développement de E^s et appliquer en x_0 .

4. Intégration.

$$J = \Delta D^{-1} = \frac{h\Delta}{\ln E} = \frac{h\Delta}{\ln(1+\Delta)} = \frac{h}{1 - \Delta/2 + \Delta^2/3 - \Delta^3/4 + \dots} = h \left(1 + \frac{1}{2}\Delta - \frac{1}{12}\Delta^2 + \dots \right)$$

$$J = h \frac{E-1}{\ln E} = h \frac{(1-\nabla)^{-1} - 1}{-\ln(1-\nabla)} = h \left(1 + \frac{1}{2}\nabla + \frac{5}{12}\nabla^2 + \frac{3}{8}\nabla^3 + \frac{251}{720}\nabla^4 + \dots \right)$$

(**Adams-Bashforth**). Sert de prédicteur dans des solveurs d'équations différentielles: $y' = f \Rightarrow y(x_1) = y(x_0) + \int_{x_0}^{x_1} f(t, y(t)) dt = y(x_0) + (Jf)(x_0)$.

$$J = h \frac{E-1}{\ln E} = h \frac{E\nabla}{-\ln(1-\nabla)} = hE \left(1 - \frac{1}{2}\nabla - \frac{1}{12}\nabla^2 - \frac{1}{24}\nabla^3 - \frac{19}{720}\nabla^4 + \dots \right)$$

(**Adams-Moulton**). Sert de correcteur.

4.1. Formules de Newton-Cotes.

Principe: on estime l'intégrale définie $\int_{x_0}^{x_m} f(x) dx$ par la valeur de l'intégrale de l'interpolant aux points équidistants x_0, x_1, \dots, x_m (formules fermées) ou x_1, x_2, \dots, x_{m-1} (formules ouvertes).

On peut par exemple exprimer $J(1 + E + \dots + E^{m-1}) = J \frac{E^m - 1}{E - 1}$ en série de puissances de Δ et s'arrêter au terme en Δ^m . Ainsi, la formule fermée pour $m = 2$ est le développement arrêté au terme en Δ^2 de $h(1 + \Delta/2 - \Delta^2/12 + \dots) \frac{(1 + \Delta)^2 - 1}{\Delta}$, ce qui donne $h(2 + 2\Delta + \Delta^2/3) = h(1 + 4E + E^2)/3$ (**formule de Simpson**). Les formules sont généralement données en termes des valeurs successives de f , comme on pourrait d'ailleurs les obtenir par intégration de la forme de Lagrange de l'interpolant:

m	Formules fermées	Formules ouvertes
1	$\frac{h(f_0 + f_1)}{2}$ trapèze	
2	$\frac{h(f_0 + 4f_1 + f_2)}{3}$ Simpson	$2hf_1$
3	$\frac{3h(f_0 + 3f_1 + 3f_2 + f_3)}{8}$ Newton 3/8, "pulcherrima"	$\frac{3h(f_1 + f_2)}{2}$
4	$\frac{2h(7f_0 + 32f_1 + 12f_2 + 32f_3 + 7f_4)}{45}$ Bode, Milne	$\frac{4h(2f_1 - f_2 + 2f_3)}{3}$

Quand m est pair, les formules fermées sont en fait exactes pour des polynômes de degré $m + 1$; $m - 1$ au lieu de $m - 2$ pour les formules ouvertes: l'interpolant de degré $m + 1$ diffère de l'interpolant de degré m par un terme constante $(x - x_0)(x - x_1) \dots (x - x_m)$ qui a une intégrale définie nulle sur $[x_0, x_m]$ si m est pair (fonction impaire de $[x - (x_0 + x_m)/2]$).

A partir des formules à 9 points, on commence à trouver des coefficients négatifs dans les formules fermées.

On utilise généralement des formules de Newton-Cotes dans des *règles composées*: l'intervalle $[a, b]$ est subdivisé en mn intervalles de longueur h , et on applique Newton-Cotes à $[x_0, \dots, x_m]$, $[x_{m+1}, \dots, x_{2m}]$, etc. Ainsi,

Règle composée des trapèzes:

$$\int_a^b f(x) dx \approx h(f_0 + 2f_1 + 2f_2 + \dots + 2f_{n-1} + f_n)/2, \quad b - a = nh,$$

Règle composée de Simpson:

$$\int_a^b f(x) dx \approx h(f_0 + 4f_1 + 2f_2 + 4f_3 + 2f_4 + 4f_5 \dots + 2f_{2n-2} + 4f_{2n-1} + f_{2n})/3, \quad b - a = 2nh.$$

5. Noyaux de Peano de règles d'intégration.

On applique (95), p. 162.

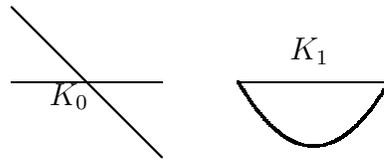
5.1. Formule du trapèze.

$$Rf = \int_{x_0}^{x_1} f(u) du - h \frac{f(x_0) + f(x_1)}{2}, \quad r = 0, m = 0 \text{ et } 1 \quad (h = b - a).$$

$$K_0(t) = \int_{x_0}^{x_1} \varphi_0(u) du - h \frac{\varphi_0(x_0) + \varphi_0(x_1)}{2} = x_1 - t - \frac{h}{2} = \frac{x_0 + x_1}{2} - t, \quad x_0 \leq t \leq x_1,$$

où $\varphi_0(u) = 0$ si $u < t$; 1 si $u > t$.

$$K_1(t) = - \int_{x_0}^t K_0(u) du = (t - x_0)(t - x_1)/2.$$



Comme K_1 est de signe constant (négatif) sur (x_0, x_1) , $Rf = Cf''(\xi)$ avec $C = \int_{x_0}^{x_1} K_1(u) du = -h^3/12$.

5.2. Formule de Simpson.

$$K_0(t) = \int_{x_0}^{x_2} \varphi_0(u) du - h \frac{\varphi_0(x_0) + 4\varphi_0(x_1) + \varphi_0(x_2)}{3} = x_2 - t - \begin{cases} 5h/3 & \text{si } x_0 < t < x_1, \\ h/3 & \text{si } x_1 < t < x_2, \end{cases}$$

soit:

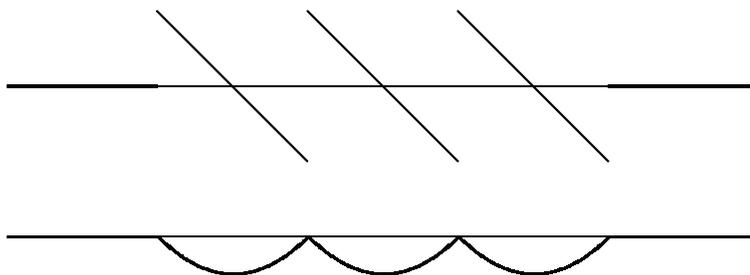
	$K_0(t)$	$K_1(t)$	$K_2(t)$	$K_3(t)$
$x_0 < t < x_1$	$x_0 + \frac{h}{3} - t$	$\frac{(t-x_0)^2}{2} - \frac{h(t-x_0)}{3}$	$\frac{-(t-x_0)^3 + h(t-x_0)^2}{6}$	$\frac{(t-x_0)^4}{24} - \frac{h(t-x_0)^3}{18}$
$x_1 < t < x_2$	$x_2 - \frac{h}{3} - t$	$\frac{(t-x_2)^2}{2} + \frac{h(t-x_2)}{3}$	$\frac{-(t-x_2)^3 - h(t-x_2)^2}{6}$	$\frac{(t-x_2)^4}{24} + \frac{h(t-x_2)^3}{18}$

Comme K_3 est de signe constant (négatif) sur (x_0, x_2) , $Rf = Cf^{iv}(\xi)$ avec $C = \int_{x_0}^{x_2} K_3(u) du = R \frac{u^4}{24} = R \frac{u-x_1)^4}{24} = \frac{h^5}{60} - h \frac{2h^4}{72} = -\frac{h^5}{90}$.

5.3. Noyaux de Peano de formules composées.

Soit une formule composée sur $[x_0, x_s] \cup [x_s, x_{2s}] \cup \dots \cup [x_{N-s}, x_N]$, où $N = qs$. Prolongeons le noyau de Peano $K_{m,i}$ sur $[x_{is}, x_{(i+1)s}]$ par $K_{m,i}(t) = 0$ si $t < x_{is}$ et $x > x_{(i+1)s}$. On a $K_{m,i}(t) = K_{m,0}(t - ish)$.

Le noyau de Peano de la formule composée est alors $K_m = \sum_{i=0}^{i=q-1} K_{m,i}$, fonction périodique sur $[x_0, x_N]$ dont la restriction à $[x_{is}, x_{(i+1)s}]$ est $K_{m,i}$.
 K_0 et K_1 de la règle des trapèzes:



etc.

Autres exemples dans Hairer [Hai, chap. 1].

5.4. La formule de Simpson avant Simpson. ¹Bien avant le développement du calcul intégral, toutes sortes de formules, vraies et fausses, ont été imaginées pour évaluer des aires et des volumes.

Johannes Kepler (1571-1630) chercha à évaluer la capacité de tonneaux de vin qu’il envisageait d’acheter à Linz en 1612 au moyen de la formule

$$\pi h \frac{r_1^2 + 4r_2^2 + r_3^2}{3},$$

où r_1 et r_3 sont les rayons des bases (souvent égaux), r_2 est le rayon à mi-hauteur, et h est la demi hauteur. Sans doute s’est-il basé sur des formules exactes connues: r_1 , r_2 et r_3 en progression arithmétique (tronc de cône); $r_1 = r_3 = 0$ (sphère). Kepler fit part de sa formule dans un ouvrage intitulé “Stereometria doliorum”, ce qui veut dire “volume des tonneaux”.

Plus tard, la formule n’apparut que comme un cas particulier des formules de Newton-Cotes (Isaac Newton, 1642-1727, en 1680; Roger Cotes² en 1711), et pas la plus intéressante: Newton préférait la règle des 3/8 qu’il appelait la “pulcherrima” (la plus belle) parce que la règle composée qu’on en tire

$$\int_{x_0}^{x_{3k}} f(x) dx \approx \frac{3h}{8} (f_0 + 3f_1 + 3f_2 + 2f_3 + \cdots + 3f_{3k-1} + f_{3k})$$

a des coefficients presque égaux. C’est bien plus tard que Thomas Simpson³ (1710-1761) se vit créditer de “sa” formule qui remporta un énorme succès, sans doute lié à l’avènement de la révolution industrielle. Mais

¹d’après des notes extraites de [Hammer], aimablement signalées par J. Meinguet.

²1682-1716, deuxième fils du Révérend Robert Cotes, recteur de Burbage (Leicestershire). Etudes et carrière à Cambridge (Trinity College). Travaux en astronomie et en physique. Il fut chargé de préparer la deuxième édition (1713) du chef-d’œuvre de Newton, *Philosophiæ naturalis principia mathematica*, dont il rédigea une préface remarquable. Ses travaux très ingénieux sur l’intégration sont rassemblés dans son opus posthume *Harmonia mensurarum* (1722). En particulier, il déduit de formules (apparemment) très différentes donnant l’aire d’un ellipsoïde de révolution, aplati ou allongé (ou haussé), cette relation pour le logarithme d’un nombre complexe de module unité: $\ln(\cos \phi + i \sin \phi) = i\phi$. (Information très aimablement communiquée par Mme Monique Van Pée).

³Thomas Simpson, Born: 20 Aug 1710 in Market Bosworth, Leicestershire, England Died: 14 May 1761 in Market Bosworth, Leicestershire, England. Simpson is best remembered for his work on interpolation and numerical methods of integration. His first job was as a weaver. At this time he taught mathematics privately and from 1737 he began to write texts on mathematics.

He also worked on probability theory and in 1740 published *The Nature and Laws of Chance*. Much of Simpson’s work in this area was based on earlier work of De Moivre.

Simpson was the most distinguished of a group of itinerant lecturers who taught in the London coffee-houses. He worked on the Theory of Errors and aimed to prove that the arithmetic mean was better than a single observation.

Simpson published the two volume work *The Doctrine and Application of Fluxions* in 1750. It contains work of Cotes. In 1754 he became editor of the *Ladies Diary*.

peut-être faut il surtout constater une force classique dans l'alignement un peu sévère et l'économie de moyens de

$$\int_{x_0}^{x_{2k}} f(x) dx \approx \frac{h}{3}(f_0 + 4f_1 + 2f_2 + 4f_3 + \cdots + 2f_{2k-2} + 4f_{2k-1} + f_{2k}),$$

et que la *pulcherrima* évoque par trop un air de valse...

6. Formule d'Euler-Maclaurin.

Polynômes de Bernoulli, formules sommatoires.

Reprenons la forme de Peano la plus simple ($m = 0$) du reste de la formule composée des trapèzes:

$$Rf = \int_{x_0}^{x_N} f(u) du - \frac{h}{2}(f_0 + 2f_1 + 2f_2 + \cdots + 2f_{N-2} + 2f_{N-1} + f_N) = \int_{x_0}^{x_N} K_0(t) f'(t) dt,$$

où K_0 est la fonction linéaire par morceaux qui vaut $x_i + h/2 - t$ entre x_i et $x_{i+1} = x_i + h$. Appelons P_1 la fonction périodique sur tout \mathbb{R} , de période 1, qui vaut $t - 1/2$ sur $(0, 1)$. On a donc $K_0(t) = -hP_1((t - x_0)/h)$. Soit $P_2/2$ une primitive de P_1 : $P_2(t) = t^2 - t +$ constante sur $(0, 1)$. P_2 est périodique car $\int_0^1 P_1(t) dt = 0$. P_2 est également *continue* sur \mathbb{R} .

On a

$$Rf = -h \int_{x_0}^{x_1} P_1((t - x_0)/h) f'(t) dt = -h^2 \int_0^N P_1(t) f'(x_0 + ht) dt = -h^2 \left[\frac{P_2}{2} f' \right]_0^N + \frac{h^3}{2} \int_0^N P_2(t) f''(x_0 + ht) dt.$$

On voit ainsi que des intégrations par parties successives livreront des puissances de plus en plus élevées de h ... Pour que l'intégrale finale garde une taille raisonnable, il faut encore que les primitives successives de P_2 restent périodiques, donc que leur intégrale définie sur $(0, 1)$ soit nulle, ce qui fixe les constantes d'intégration en suspens:

Définitions. Les fonctions périodiques de Bernoulli ⁴ P_0, P_1, P_2, \dots sont de période 1, $P_0 = 1$, P_n/n est la primitive périodique de P_{n-1} , c'est-à-dire telle que $P'_n = nP_{n-1}$ et

$$\int_0^1 P_n(t) dt = 0 \text{ quand } n \geq 1;$$

les polynômes de Bernoulli B_n coïncident avec les fonctions P_n sur $(0, 1)$;

les nombres de Bernoulli sont $B_n = B_n(0)$.

The following description of Simpson by Charles Hutton (made 35 years after Simpson's death) is interesting *It has been said that Mr Simpson frequented low company, with whom he used to guzzle porter and gin: but it must be observed that the misconduct of his family put it out of his power to keep the company of gentlemen, as well as to procure better liquor.*

It would be fair to note that others described Simpson's conduct as irreproachable.

Simpson also worked on the problem of the minimum length of path that must be used to join three points. Tell me about Simpson's work on minimal paths

Thomas Simpson was elected to the Royal Society of London in 1745.

cf. <http://www-history.mcs.st-and.ac.uk/history/Mathematicians/Simpson.html>

⁴Il s'agit de Jakob Bernoulli (1654-1705), frère de Johann (1667-1748) et oncle de Daniel (1700-1782).

Par primitivations successives, on trouve les premiers échantillons

$$B_0(x) \equiv 1; B_1(x) = x - \frac{1}{2}; B_2(x) = x^2 - x + \frac{1}{6}; B_3(x) = x^3 - \frac{3x^2}{2} + \frac{x}{2}; B_4(x) = x^4 - 2x^3 + x^2 - \frac{1}{30}; \dots$$

Théorème (formule d'Euler-Maclaurin). Soit $f \in \mathcal{C}^{2m+2}[a, b]$ et $h = (b - a)/N$, alors

$$\int_a^b f(t) dt = \frac{h}{2}(f_0 + 2f_1 + 2f_2 + \dots + 2f_{N-2} + 2f_{N-1} + f_N) - \sum_{k=1}^m \frac{B_{2k}}{(2k)!} h^{2k} [f^{(2k-1)}(b) - f^{(2k-1)}(a)] + \varepsilon_{2m+2} \quad (99)$$

avec $\varepsilon_{2m+2} = -h^{2m+2}(b - a) \frac{B_{2m+2}}{(2m+2)!} f^{(2m+2)}(\xi) = -h^{2m+2} \frac{B_{2m+2}}{(2m+2)!} \sum_{j=0}^{N-1} f^{(2m+2)}(\xi_j)$, où $\xi \in [a, b]$, $x_j \leq \xi_j \leq x_{j+1}$ ($f_j = f(x_j) = f(a + jh)$, $j = 0, 1, \dots, N$).

En effet, nous sommes arrivés à $Rf = \int_a^b f(t) dt -$ somme des trapèzes $= -h^2 \left[\frac{P_2}{2} f' \right]_0^N + h^3 \int_0^N \frac{P_2(t)}{2} f''(a + ht) dt$. Effectuons encore $2m$ intégrations par parties: on obtient des termes aux limites $(-1)^{j-1} h^j \left[\frac{P_j(t)}{j!} f^{(j-1)}(a + ht) \right]_{t=0}^{t=N}$ pour $j = 2, 3, \dots, 2m + 2$ et une intégrale finale $h^{2m+3} \int_0^N \frac{P_{2m+2}(t)}{(2m+2)!} f^{(2m+2)}(t) dt$. Comme P_j est périodique continue pour $j > 1$, $P_j(N) = P_j(0) = B_j$, et on verra que $B_j = 0$ si j est impair ≥ 3 : on trouve bien la somme figurant dans (99), en fait, on trouve un terme en plus, le dernier terme étant $-h^{2m+2} (B_{2m+2} / ((2m+2)!)) (f^{(2m+1)}(b) - f^{(2m+1)}(a))$.

Quant à l'intégrale finale, on l'écrit

$$h^{2m+2} \int_a^b \left[\frac{P_{2m+2}((t-a)/h) - B_{2m+2}}{(2m+2)!} + \frac{B_{2m+2}}{(2m+2)!} \right] f^{(2m+2)}(t) dt = \varepsilon_{2m+2} + h^{2m+2} \frac{B_{2m+2}}{(2m+2)!} \int_a^b f^{(2m+2)}(t) dt$$

de façon à bénéficier d'une fonction $P_{2m+2} - B_{2m+2}$ de signe constant (on verra aussi plus tard que $|P_{2j}(t)| \leq |P_{2j}(0) = B_{2j}|$), d'où une première contribution $h^{2m+2} f^{(2m+2)}(\xi)$ fois l'intégrale sur (a, b) de $[P_{2m+2}((t-a)/h) - B_{2m+2}] / ((2m+2)!)$ qui se réduit à l'intégrale de la constante $-B_{2m+2} / ((2m+2)!)$ puisque P_{2j} a une intégrale définie nulle dès que $j \geq 1$. Comme $P_{2m+2}((t-a)/h)$ est périodique de période h , on a la deuxième expression de ε_{2m+2} en sommant les intégrales sur des intervalles de longueur h .

Enfin, l'intégrale de la constante B_{2m+2} fois $f^{(2m+2)}(t)$ donne évidemment B_{2m+2} fois $f^{(2m+1)}(b) - f^{(2m+1)}(a)$, d'où la somme s'arrêtant à $k = m$ dans (99). \square

En termes d'opérateurs aux différences, il faut exprimer $Rf = J(1 + E + \dots + E^{N-1}) - h(1 + 2E + 2E^2 + \dots + 2E^{N-1} + E^N)/2$ appliqué à f en $a = x_0$.

On obtient $J \frac{E^N - 1}{E - 1} - h \left(\frac{E^N - 1}{E - 1} + \frac{E^N - 1}{2} \right) = (E^N - 1) \left(\frac{J}{E - 1} - \frac{h}{E - 1} - \frac{h}{2} \right)$.

En exprimant E dans la dernière parenthèse, tantôt en fonction de Δ , tantôt en fonction de ∇ , on a

$$\frac{J - h}{E - 1} - \frac{h}{2} = \frac{h}{\ln(1 + \Delta)} - \frac{h}{\Delta} - \frac{h}{2} = -\frac{h\Delta}{12} + \frac{h\Delta^2}{24} - \frac{19h\Delta^3}{720} + \frac{3h\Delta^4}{160} - \dots$$

et

$$\frac{J - h}{E - 1} - \frac{h}{2} = -\frac{h}{\ln(1 - \nabla)} - \frac{h}{\nabla} + \frac{h}{2} = -\frac{h\nabla}{12} - \frac{h\nabla^2}{24} - \frac{19h\nabla^3}{720} - \frac{3h\nabla^4}{160} - \dots$$

d'où

$$Rf = -\frac{h}{12}((\nabla f)(x_N) - (\Delta f)(x_0)) - \frac{h}{24}((\nabla^2 f)(x_N) + (\Delta^2 f)(x_0)) - \frac{19h}{720}((\nabla^3 f)(x_N) - (\Delta^3 f)(x_0)) \\ - \frac{3h}{160}((\nabla^4 f)(x_N) + (\Delta^4 f)(x_0)) - \dots$$

(formule de Gregory).

6.1. Identités des nombres et polynômes de Bernoulli.

Si on veut retrouver (99) par les opérateurs aux différences, il faut exprimer

$(E^N - 1) \left(\frac{J}{E - 1} - \frac{h}{E - 1} - \frac{h}{2} \right)$ en fonction de D par $E = e^{hD}$ et $J = \Delta D^{-1} = (E - 1)D^{-1}$:
 $Rf = (E^N - 1)(D^{-1} - h/(e^{hD} - 1) - h/2)$ appliqué à f en $a = x_0$.

Considérons le développement $\frac{x}{e^x - 1} = \sum_{j=0}^{\infty} X_j x^j$, alors

$$Rf = -[(E^N - 1) \sum_{j=2}^{\infty} X_j h^j D^{j-1}]f(x_0) = -\sum_{j=2}^{\infty} X_j h^j [f^{(j-1)}(b) - f^{(j-1)}(a)],$$

en ayant rapidement vérifié $X_0 = 1, X_1 = -1/2$. On retrouve bien (99) et on soupçonne que $X_j = B_j/j!$. En effet:

Fonction génératrice des polynômes de Bernoulli.

$$\sum_{j=0}^{\infty} \frac{B_n(x)t^n}{n!} = \frac{te^{xt}}{e^t - 1},$$

la série converge quand $|t| < 2\pi$.

En effet, soit $G(x; t) := \sum_{j=0}^{\infty} \frac{B_n(x)t^n}{n!}$. Comme $B'_n = nB_{n-1}$,⁵

$\partial G(x; t)/\partial x = \sum_{n=1}^{\infty} B_{n-1}(x)t^n/(n-1)! = tG(x; t)$, d'où $G(x; t) = K(t)e^{xt}$. Comme $B_n(1) = B_n(0)$ quand $n > 1$ (et aussi, bien sûr, quand $n = 0$),

$$G(1; t) - G(0; t) = \begin{cases} [B_1(1) - B_1(0)]t & = t \quad \text{on sait que } B_1(x) = x - 1/2 \\ K(t) (e^t - 1) \end{cases}$$

⁵Les polynômes de Bernoulli forment donc une suite d'Appell, comme les polynômes d'Hermite.

donc, $K(t) = t/(e^t - 1)$. Les pôles sont en $t = 2k\pi i$, $k = \pm 1, \pm 2, \dots$ □

Montrons que $B_j = 0$ quand j est impair > 1 :

$$\sum_2^{\infty} B_j t^j / j! = \frac{t}{e^t - 1} + \frac{t}{2} - 1 = \frac{t e^t + 1}{2 e^t - 1} - 1 = \frac{t}{2} \operatorname{cotangh} \frac{t}{2} - 1$$

est une fonction paire de t .

Exercice. Montrez que $\operatorname{tg} x = \sum_{n=0}^{\infty} (-1)^n \frac{4^{n+1}(4^{n+1} - 1) B_{2n+2}}{(2n+2)!} x^{2n+1}$.

(Utilisez $\operatorname{tg} x = \cotg x - 2 \cotg(2x)$.)

Coefficients des polynômes de Bernoulli: d'après $B'_n = n B_{n-1}$,

$$B_n(x) = \sum_{k=0}^n \frac{B_n^{(k)}(0)}{k!} x^k = \sum_{k=0}^n \frac{n(n-1)\dots(n-k+1) B_{n-k}(0)}{k!} x^k = \sum_{k=0}^n \binom{n}{k} B_{n-k} x^k.$$

Calcul symbolique. Des identités de nombres et polynômes de Bernoulli s'obtiennent à partir d'un symbole \mathbf{B} dont les puissances \mathbf{B}^k sont remplacées en fin de compte par les nombres de Bernoulli B_k . Ainsi:

$$B_n(x) = (\mathbf{B} + x)^n.$$

On appelle *calcul umbral* cette façon de faire, cf. S. Roman, G.-C. Rota, The umbral calculus, *Adv. Math.* **27** (1978) 95–188, G.-C. Rota, B.D. Taylor, The classical umbral calculus, *SIAM J. Math. An.* **25** (1994) 694–711, A. Di Bucchianico, *Probabilistic and analytical aspects of the umbral calculus*, Centrum voor Wiskunde and Informatica Tract **119**, Amsterdam, 1997. Voir aussi

sci.math #144246 (0 + 1059 more)

From: wgd@zurich.ai.mit.edu (Bill Dubuque)

Newsgroups: sci.math

[2] Bernoulli numbers everywhere: Umbral Calculus [was: Re: n:th Bernoulli number]

Date: 27 Jun 96 18:53:53

Organization: M.I.T. Artificial Intelligence Lab.

Message-ID: <WGD.96Jun27185353@berne.ai.mit.edu>

NNTP-Posting-Host: berne.ai.mit.edu

In-reply-to: bruck@pacificnet.net's message of Tue, 25 Jun 1996 18:51:52 -0700

Digressing somewhat, it is surprising just how often Bernoulli numbers and polynomials arise in diverse contexts. For example, Lang sketches how they arise in topology and algebraic geometry around Riemann-Roch theorems, and in analytic and algebraic number theory around zeta functions and modular forms (see the thread of exercises beginning with #21 p. 217, end of Chap. IV in Lang's *Algebra*, 3rd Ed.) One way of understanding this ubiquity comes from the viewpoint of Hopf algebras and coalgebras, or, equivalently, the Umbral Calculus. The latter provides a calculus of adjoints that serves to systematically derive and classify almost all of the classical combinatorial identities for polynomial sequences (e.g the sequences of Abel, Appel, Bell, Bernoulli, Bessel, Boole, Boas-Buck, Euler, Gould, Hermite, Laguerre, Mahler, Meixner, Mittag-Leffler, Mott, Poisson-Charlier, Sheffer, Stirling, etc), along with associated identities (generating functions, expansions, duplication formulas, recurrences, inversions, Rodrigues formula, etc, e.g. the Euler-Maclaurin expansion, Boole's Summation formula, Newton interpolation, Gregory integration, Vandermonde convolution). For example almost all of the identities in Riordan's classic book "Combinatorial Identities" can be systematically derived and classified via the Umbral Calculus. This is a powerful and useful theory that deserves to be better known. Since there are now good expositions available, there's no longer any good reason not to have a peek. I recommend starting with the following online survey [included in ascii below those Web challenged].

<http://www.win.tue.nl/win/math/bs/statistics/bucchianico/hypersurvey/hypersurvey.html>

See also the RotaFest page at <http://www-math.mit.edu/~loeb/rotafest.html>

-Bill

Note: math formulas are missing from this ascii version, see the URL above for them.

Récurrence des nombres de Bernoulli. Développons $B_{n+1}(1) = B_{n+1}(0)$:

$$\sum_{k=1}^{n+1} \binom{n+1}{k} B_{n+1-k} = 0.$$

On obtient ainsi quelques nombres de Bernoulli⁶ non nuls:

B_0	B_1	B_2	B_4	B_6	B_8	B_{10}	B_{12}	B_{14}	B_{16}	B_{18}	B_{20}
1	$-\frac{1}{2}$	$\frac{1}{6}$	$-\frac{1}{30}$	$\frac{1}{42}$	$-\frac{1}{30}$	$\frac{5}{66}$	$-\frac{691}{2730}$	$\frac{7}{6}$	$-\frac{3617}{510}$	$\frac{43867}{798}$	$-\frac{174611}{330}$

cf. S. Plouffe (editor): The first 498 Bernoulli Numbers, (2001), <http://ia300234.us.archive.org/>

Sommes de puissances consécutives. Appliquons la formule d'Euler-Maclaurin à $f(x) = x^r$, r entier ≥ 0 , avec $a = 0$, $b = N + 1$, $h = 1$:

$$(N+1)^{r+1}/(r+1) = 1 + 2^r + \dots + N^r + \frac{(N+1)^r}{2} - \sum_{k=2}^{r+1} \frac{B_k}{k!} r(r-1)\dots(r-k+2)(N+1)^{r-k+1},$$

on arrive à

$$1 + 2^r + \dots + N^r = [B_{r+1}(N+1) - B_{r+1}(0)]/(r+1).$$

Plus généralement, $B_n(x+1) - B_n(x) = nx^{n-1}$: par la fonction génératrice, $B_n(x+1) - B_n(x)$ est $n!$ fois le coefficient de t^n de $t(e^{(x+1)t} - e^{xt})/(e^t - 1) = te^{xt}$, ce qui donne bien nx^{n-1} .

Séries de Fourier des extensions périodiques des polynômes de Bernoulli. Partant de

$$P_1(x) = \sum_{k=-\infty}^{\infty} \int_0^1 (t-1/2)e^{-2\pi ikt} dt e^{2\pi ikx} = \frac{-1}{2\pi i} \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} k^{-1} e^{2\pi ikx},$$

on intègre $n-1$ fois:

$$P_n(x) = \frac{-n!}{(2\pi i)^n} \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} k^{-n} e^{2\pi ikx}.$$

Quand n est pair,

$$P_n(x) = \frac{-2(-1)^{n/2}n!}{(2\pi)^n} \sum_{k=1}^{\infty} k^{-n} \cos(2\pi kx),$$

ce qui montre que $|P_n(x)| \leq |P_n(0)|$ (tous les coefficients de la série de Fourier sont de même signe).

⁶Attention, certains auteurs (Dwight, par exemple) utilisent une autre convention où $(-1)^{k-1}B_{2k}$ est noté B_k .

On a donc $B_{2n} = \frac{2(-1)^{n+1}(2n)!}{(2\pi)^{2n}} \sum_{k=1}^{\infty} k^{-2n}$, $n = 1, 2, \dots$, ainsi:

$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}, \quad \sum_{k=1}^{\infty} \frac{1}{k^4} = \frac{\pi^4}{90}, \quad \sum_{k=1}^{\infty} \frac{1}{k^6} = \frac{\pi^6}{945}, \quad \text{etc.}$$

Obtention directe de $\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$.

Le polynôme $\frac{(x+i)^{2n+1} - (x-i)^{2n+1}}{2i} = c_0x^{2n} + c_2x^{2n-2} + \dots = (2n+1)x^{2n} - (2n+1)(2n)(2n-1)x^{2n-2}/6 + \dots$

a $2n$ zéros distincts $z_k = \cotg \frac{k\pi}{2n+1}$, $k = 0, 1, \dots, 2n$ symétriquement disposés autour de 0: $z_k = -z_{2n+1-k}$.

La somme des carrés des zéros est $(c_1/c_0)^2 - 2c_2/c_0$, ici $2n(2n-1)/3$, donc

$$\frac{n(2n-1)}{3} = \sum_1^n \cotg^2 \frac{k\pi}{2n+1} < \frac{(2n+1)^2}{\pi^2} \sum_1^n \frac{1}{k^2} < \sum_1^n \operatorname{cosec}^2 \frac{k\pi}{2n+1} = \frac{n(2n+2)}{3}.$$

Cf. H. Tsumura, $\zeta(2m)$, *Amer. Math. Monthly* **111** (2004) 430–431.

D.H. Lehmer (A new approach to Bernoulli polynomials, *Amer. Math. Monthly* **95** (1988) 905–911) distingue 5 façons d’aborder les nombres et polynômes de Bernoulli:

Somme de puissances $\sum_0^{m-1} k^{n-1} = (B_n(m) - B_n(0))/n$ (Jac. Bernoulli, avant 1705),

Fonction génératrice $\sum_0^{\infty} B_n(x)t^n/n! = te^{xt}/(e^t - 1)$ (Euler, 1738),

Suites d’Appell $B_{n-1} = B'_n/n$ (Appell, 1832),

Séries de Fourier $B_n(x) = -n!/(2\pi i)^n \sum_{k=-\infty}^{\infty} k^{-n} e^{2\pi i k x}$ (Hürwitz, 1890),

Calcul umbral $B_n(x) = (B + x)^n$ (Lucas, 1891),

auxquelles il ajoute une sixième, basée sur une formule de multiplication de Raabe (1851): $B_n(mx) = m^{n-1} \sum_{k=0}^{m-1} B_n(x + k/m)$.

Nombres et polynômes d’Euler.

$$\frac{2e^{xt}}{e^t + 1} = \sum_0^{\infty} E_n(x) \frac{t^n}{n!}, \quad E_n = 2^n E_n(1/2).$$

$E_0 \quad E_2 \quad E_4 \quad E_6 \quad E_8 \quad E_{10} \quad E_{12} \quad E_{14}$

1 -1 5 -61 1385 -50521 2702765 -199360981

On a $\frac{1}{\cos x} = \sum_{n=0}^{\infty} (-1)^n \frac{E_{2n} x^{2n}}{(2n)!}$.

6.2. Schéma d’intégration de Romberg.

Comme $R_h f := \int_{x_0}^{x_N} f(u) du - \frac{h}{2}(f_0 + 2f_1 + 2f_2 + \dots + 2f_{N-2} + 2f_{N-1} + f_N)$ a, d’après (99),

un développement en puissances *paires* de h , on applique l’extrapolation de Richardson partant des $c_{i,0} = R_{h_0/2^i}$, puis

$$c_{i,k} = \frac{(x - x_i)c_{i+1,k-1}(x) - (x - x_{i+k})c_{i,k-1}}{x_{i+k} - x_i}, \quad k > 0.$$

avec $x = 0$ et $x_i = h_0/4^i$:

$$c_{i,k} = \frac{4^k c_{i+1,k-1}(x) - c_{i,k-1}}{4^k - 1}, \quad k > 0.$$

6.3. Formule d'Euler-Maclaurin en tant que formule sommatoire.

Reprenons (99) en exprimant la somme:

$$f_0 + f_1 + f_2 + \cdots + f_{N-1} = h^{-1} \int_a^b f(t) dt + \frac{f_0 - f_N}{2} + \sum_{k=1}^m \frac{B_{2k}}{(2k)!} h^{2k-1} [f^{(2k-1)}(b) - f^{(2k-1)}(a)] - \varepsilon_{2m+2}/h, \quad (100)$$

où $\varepsilon_{2m+2} = -h^{2m+2} \frac{B_{2m+2}}{(2m+2)!} \sum_{j=0}^{N-1} f^{(2m+2)}(\xi_j)$, avec $\xi_j \in [x_j, x_{j+1}]$ ($f_j = f(x_j) = f(a+jh)$, $j = 0, 1, \dots, N$).

Par un intéressant renversement de valeurs, on peut estimer une somme d'un grand nombre de valeurs fonctionnelles en partant d'une intégrale!

Nombres harmoniques, constante d'Euler.

Le $M^{\text{ème}}$ nombre harmonique est le nombre rationnel $H_M = 1 + 1/2 + \cdots + 1/M$. Soit M_0 grand $< M$, on a $H_M = H_{M_0} + 1/(M_0 + 1) + \cdots + 1/M$. Par la formule précédente, avec $N = M - M_0$, $f(x) = 1/(x + M_0)$ et $h = 1$, $H_M = H_{M_0} - 1/M_0 + 1/M + \ln(M/M_0) + (1/M_0 -$

$$1/M)/2 - \sum_{k=1}^m \frac{B_{2k}}{2k} \left(\frac{1}{M_0^{2k}} - \frac{1}{M^{2k}} \right) - \varepsilon_{2m+2}$$

donc,

$$H_M - \ln M - \frac{1}{2M} + \sum_{k=1}^{2m} \frac{B_{2k}}{2kM^{2k}} = H_{M_0} - \ln M_0 - \frac{1}{2M_0} + \sum_{k=1}^{2m} \frac{B_{2k}}{2kM_0^{2k}} - \varepsilon_{2m+2},$$

avec $|\varepsilon_{2m+2}| = |B_{2m+2}| \sum_{j=0}^{N-1} (M_0 + j + \theta_j)^{-2m-3} < \frac{|B_{2m+2}|}{(2m+2)(M_0-1)^{2m+2}}$, ce qui montre que $\{H_M - \ln M\}$ est une suite de Cauchy de limite appelée γ , **constante d'Euler**. Avec $M = \infty$:

$$\gamma = H_{M_0} - \ln M_0 - \frac{1}{2M_0} + \sum_{k=1}^{2m} \frac{B_{2k}}{2kM_0^{2k}} - \varepsilon_{2m+2},$$

ce qui permet de calculer γ efficacement:

Dear Simon,

I send you Euler's constant to 1000000 digits. The algorithm used is the one presented in the paper of McMillan and Brent [1] (the second variant) accelerated by binary splitting [2], [3]. This calculation verified the previous 500000-digits record to 499927 digits. Using the 500000-digit value of gamma and a program by H.J.J. te Riele, CWI Amsterdam [4], I was also able to compute 470006 partial quotients of the continued fraction of gamma. Computing the 470006-th convergent results in the following [5]

Theorem: If Euler's constant is a rational number P/Q , for integers P, Q then $|Q| > 10^{242080}$

* Acknowledgements *

The latest calculation would have been possible without the help and motivation from H.J.J. te Riele, Richard Brent and my colleague Bruno Haible.

Best

Thomas Papanikolaou

References:

- [1] R. P. Brent and E. M. McMillan, Some new algorithms for high-precision computation of Euler's constant, *Math. Comp.* **34** (1980) 305-312.
- [2] Bruno Haible, Thomas Papanikolaou, Fast multiprecision evaluation of series of rational numbers, Technical Report No. TI-7/97, 18.03.1997. Available via <http://www.informatik.th-darmstadt.de/TI/Veroeffentlichung>
- [3] Jonathan M. Borwein and Peter B. Borwein, *Pi and the AGM*, Wiley, 1987.
- [4] Richard P. Brent, Alfred J. van der Poorten and Herman J.J. te Riele, A comparative study of algorithms for computing continued fractions of algebraic numbers, pp. 37-49 in: H. Cohen (editor), *Algorithmic Number Theory: Second International Symposium, ANTS-II*, Talence, France, 18-23.05.1996, Springer, Berlin etc., 1996.
- [5] Thomas Papanikolaou, Entwurf und Entwicklung einer objektorientierten Bibliothek für algorithmische Zahlentheorie, PhD Thesis, to appear.
- [6] Steve Finch, <http://www.mathsoft.com/asolve/constant/euler/euler.html>

Euler's constant to 1000000 digits computed on Mars 24, 1997 on a Sun Sparc Ultra machine with 167 Mhz and 256 MB RAM in 42 hours 27 hsec.

The algorithm was implemented using the LiDIA library for computational number theory and it will be part of the multiprecision floating-point arithmetic of the package in release 1.4. LiDIA is available from

<ftp://ftp.informatik.th-darmstadt.de/pub/TI/systems/LiDIA> , <http://www.informatik.th-darmstadt.de/>
 LiDIA's kernel used Bruno Haible's CLN package, which is available via anonymous FTP from <ftp://ma2s2.mathematik.uni-karlsruhe.de/pub/gnu/cln.tar.z>
 This package contains the first implementation of a binary splitting device for rational series.

Here is the output of the program:

Calculating Euler's constant to 1000000 decimals

Time required: 42 hour 27 hsec

Euler =
 0.5772156649015328606065120900824024310421593359399235988057672348848677267776\
 646709369470632917467495146314472498070824809605040144865428362241739976449235\
 \

etc. cf. <http://www.lacim.uqam.ca/pi>

Grandes factorielles. Formule de Stirling. On reprend le raisonnement précédent avec $\ln M$ au lieu de $1/M$:

$$\ln M! - M \ln M + M - \frac{\ln M}{2} - \sum_{k=1}^{2m} \frac{B_{2k}}{2k(2k-1)M^{2k-1}} = \ln M_0! - M_0 \ln M_0 + M_0 - \frac{\ln M_0}{2} - \sum_{k=1}^{2m} \frac{B_{2k}}{2k(2k-1)M_0^{2k-1}} - \varepsilon_{2m+2},$$

avec $|\varepsilon_{2m+2}| = \frac{|B_{2m+2}|}{2m+2} \sum_{j=0}^{N-1} (M_0 + j + \theta_j)^{-2m-2} < \frac{|B_{2m+2}|}{(2m+2)(2m+1)(M_0-1)^{2m+1}}$, ce qui montre que $\{\ln M! - (M + 1/2) \ln M + M\}$ est une suite de Cauchy. La limite $\ln \sqrt{2\pi}$ s'établit par examen d'inégalités portant sur $\int_{-\pi/2}^{\pi/2} \cos^{2M} \theta d\theta = \frac{\pi}{4^M} \binom{2M}{M}$. Donc,

$$M! = \sqrt{2\pi} M^{M+1/2} e^{-M} \exp \left\{ \sum_{k=1}^{2m} \frac{B_{2k}}{2k(2k-1)M^{2k-1}} + \varepsilon_{2m+2} \right\}.$$

Fonction zeta de Riemann. $\zeta(s) = \sum_1^\infty \frac{1}{n^s}$ converge absolument seulement si $\text{Re } s > 1$.

$\zeta(s) = \sum_1^{N-1} \frac{1}{n^s} + \sum_N^\infty \frac{1}{n^s}$. On applique (100) à la deuxième série, avec $f(x) = x^{-s}$, $a = N$, $b = \infty$ et $h = 1$. On trouve

$$\zeta(s) = \sum_1^{N-1} \frac{1}{n^s} + \frac{1}{2N^s} + \frac{1}{(s-1)N^{s-1}} + \sum_{k=1}^m \frac{B_{2k}}{(2k)!} \frac{s(s+1)(s+2)\dots(s+2k-2)}{N^{s+2k-1}} - \varepsilon_{2m+2},$$

où $\varepsilon_{2m+2} = -\frac{B_{2m+2}}{(2m+2)!} s(s+1)\dots(s+2m+1) \sum_{j=N}^\infty \frac{1}{\xi_j^{s+2m+2}}$, avec $\xi_j \in [j, j+1]$.

Cette dernière série converge, et est aussi petite que l'on veut pourvu que N soit assez grand, dès que $\text{Re } s > 1 - 2m$. On peut ainsi définir, et calculer, $\zeta(s)$ dans tout $\mathbb{C} \setminus \{1\}$.

Autres formules sommatoires.

Plana:

$$\sum_{n=0}^\infty f(n) = \frac{1}{2}f(0) + \int_0^\infty f(x) dx + i \int_0^\infty \frac{f(iy) - f(-iy)}{e^{2\pi y} - 1} dy$$

(P. Henrici, *Applied and Computational Complex Analysis I*, Wiley, 1974, p. 274).

Boole:

$$\sum_{n=0}^{n=N-1} (-1)^n f(a + sh + nh) = \sum_{k=0}^{k=m-1} \frac{h^k}{2k!} E_k(s) [f^{(k)}(a) - (-1)^N f^{(k)}(b)] + \varepsilon_m.$$

(N.E. Nörlund, *Vorlesungen über Differenzenrechnung*, chap. 2; J.M. Borwein, P.B. Borwein, K. Dilcher: Pi, Euler numbers, and asymptotic expansions, *American Math. Monthly* **96** (1989) 681-687.).

Poisson:

$$\sum_{k=-\infty}^{\infty} F(k\eta) = \frac{1}{\eta} \sum_{n=-\infty}^{\infty} G(2\pi n/\eta),$$

où G est la transformée de Fourier de F : $G(\omega) = \int_{-\infty}^{\infty} F(\sigma)e^{-i\omega\sigma} d\sigma$

(P. Henrici, *Applied and Computational Complex Analysis II*, Wiley, 1977, p. 270).

Règle de la tangente hyperbolique:

$$\int_{-1}^1 f(x) dx = \int_{-\infty}^{\infty} f(\operatorname{tgh} t) \frac{dt}{\cosh^2 t} \approx h \sum_{-\infty}^{\infty} \frac{f(\operatorname{tgh} nh)}{\cosh^2 nh}.$$

Nombreuses propriétés curieuses⁷



⁷C. Schwartz, Numerical integration of analytic functions, *J. Comput. Phys.* **4** (1969) 19-29. S. Haber, The tanh rule for numerical integration, *SIAM J. Numer. Anal.* **14** (1977) 668-685. M. Mori, Quadrature formulas obtained by variable transformation and the DE-rule, *J. Comp. Appl. Math.* **12& 13** (1985) 119-130.

CHAPITRE 6

Récapitulation.

Définition de l'approximation	Principe directeur (Existence, unicité caractérisation)	Méthode d'obtention	Estimation d'erreur, convergence	Développements applications algorithmes
Meilleure approx. en norme uniforme $\min \ f - p\ _\infty$	Théorème d'équi-oscillation	Algorithme d'échange, bonne approx. $\sum' c_k(f)T_k$	de La Vallée Poussin Weierstrass	Propriétés des T_n , \perp discrète, réarrangement, récurrence des T_n
Meilleure approximation en moyenne quadratique $\ f - p\ _2$ minimum	$p =$ projection orthogonale de f dans V	cas polynomial: polynômes orthogonaux. Cas trigonométrique: séries de Fourier	suites totales, max. espace de Hilbert vitesse de décroissance coeff. Fourier	Propriétés des polynômes orthogonaux. FFT
Interpolation $\lambda_i(p) = \lambda_i(f)$ (par ex., $p(a_i) = f(a_i)$)	Système d'équations linéaires.	Lagrange, Newton, Neville, Hermite Newton-Gregory	Peano	\int de Gauss Newton-Cotes, Bickley Euler-Maclaurin

CHAPITRE 7

Appendices: alphabets grec, cyrillique, petit dico, index.

1. Alphabets

<i>A</i>	α	alpha
<i>B</i>	β	beta
Γ	γ	gamma
Δ	δ	delta
<i>E</i>	ϵ	epsilon
<i>Z</i>	ζ	zeta
<i>H</i>	η	eta
Θ	θ, ϑ	theta
<i>I</i>	ι	iota
<i>K</i>	κ	kappa
Λ	λ	lambda
<i>M</i>	μ	mu
<i>N</i>	ν	nu
Ξ	ξ	ksi
<i>O</i>	o	omicron
Π	π	pi
<i>R</i>	ρ	rho
Σ	σ, ς	sigma
<i>T</i>	τ	tau
Υ	υ	upsilon
Φ	ϕ, φ	phi
χ	χ	khi
Ψ	ψ	psi
Ω	ω	omega

А	а	A
Б	б	B
В	в	V
Г	г	G
Д	д	D
Е	е	E
Ж	ж	zh
З	з	Z
И	и	I
Ј	ј	J
К	к	K
Л	л	L
М	м	M
Н	н	N
О	о	O
П	п	P
Р	р	R
С	с	S
Т	т	T
У	у	ou
Ф	ф	F
Ч	ч	tch
Ш	ш	ch
Щ	щ	chtch
Ы	ы	Y
Э	э	è
Ю	ю	you
Я	я	ya

2. Petit dico mathematical English → français mathématique.

Y compris des expressions jugées utiles lors du *Mathematics Seminar* MATH 2900.

Voir aussi dictionnaires mathématiques français-anglais et anglais-français dans “EPS MAG DE MATHS” <http://perso.wanadoo.fr/eps/>

Hauchecorne Bertrand, Shaw Adrian: *Lexique bilingue du vocabulaire mathématique*, ELLIPSES - Edition Marketing S.A., Paris

analysis: analyse	number: nombre
to approximate: approcher	numeric, numerical: numérique(s)
approximation of functions	odd: impair
approximation to a function	outcome: résultat
on the assumption that: dans l’hypothèse où	polynomial: polynôme, polynomial
ball: boule	positive: strictement positif(ve)
beyond: au delà	principle: principe
bound: borne	provided: pourvu que
bounded: borné(e)	Q.E.D. : C.Q.F.D.
Chebyshev: Tchebycheff	quotient space (factor space): espace quotient
decreasing: strictement décroissant(e)	radius, radii: rayon, rayons
digit: chiffre	rational: rationnel(le)
eigenvalue, vector: valeur, vecteur propre	remainder: reste
even: pair	reproducing: reproduisant, régénérateur
ever: toujours, indéfiniment	root: racine (voir zero)
field: corps, champ	rule: règle
for all: pour tout(e)	sample: échantillon
function: fonction	scalar: scalaire
generating function: fonction génératrice	sequence: suite
generating sequence: suite génératrice	series: série(s)
increasing: strictement croissant(e)	small: petit
induced map: application induite	space: espace
kernel: noyau	spaced out: échelonné
key idea: idée principale	spanned by: engendré par
large: grand	square: carré
least squares: moindres carrés	stable under: stable par
least squares approximation:	to state: déclarer
approxim. en moyenne quadratique	such that: tel(le) que
less than: strictement inférieur à	symmetric: symétrique
link: lien	tailor: tailleur
map, mapping: application (math.)	Taylor: Taylor
negative: strictement négatif(ve)	unbounded: non borné(e)
non decreasing: croissant(e)	unless: à moins que
non increasing: décroissant(e)	until: jusqu’à ce que
non negative: positif(ve)	to vanish: s’annuler
non positive: négatif(ve)	vector space: espace vectoriel
norm: norme	zero: zéro (d’une fonction)

Index

- absolument continue, fonction, 75
- ACM, 10
- affine, espace, 23
- AGM, 16
- aliasing, 64
- Apéry, 18
- Approximation rationnelle, 67
- autoadjoint (form.), 101

- B-splines, 142
- base de Lagrange, 147
- base duale, 51, 147
- Beltrami, 109
- Ben-Naoum, K., 17
- Bernoulli, nombres et polynômes de, 174
- Bernstein, 42, 135
- Bernstein, polynômes de, 132
- Bessel, inégalité de, 129
- Bezier, 141
- Bickley, 170
- binomiale, loi, 132
- biorthogonalité, 52, 147
- Boole, formule sommatoire de, 182
- Brezinski, 153

- Carathéodory-Fejér, 68
- Christoffel-Darboux, 47, 100
- classiques, polynômes orthogonaux, 102, 107
- Clenshaw-Curtis, règle de, 161
- convexe, 23
- CORDIC, 14
- Cotes, R., 173

- Darboux-Christoffel, 47, 100
- de La Vallée Poussin, 31, 60, 70
- definie positive, 73, 77
- Delsarte, P., 107
- différences divisées, 150
- Dirichlet, 56
- distance, 19
- duale, base, 51

- échange, algorithme d', 33

- electrostatique, 105
- Equations du 3ème degré, 47
- equations normales, 79, 115
- equioscillation, 28
- Euler, constante d', 180
- Euler, nombres et polynômes d', 179
- Euler-Maclaurin, formule d', 175

- Fejer, 136
- FFT, 124
- Fourier, 56, 118
- Fourier, séries de, 56
- fraction continue, 17, 98
- Frobenius, 67
- Frutiger, 142

- Gamma, fonction, 93
- Gegenbauer, polynômes de, 107
- generatrice, fonction, 46
- Genin, Y., 107
- Genocchi, 155
- Gram, matrice de, 77
- Gram-Schmidt, 80, 87

- Hölder, normes de, 20
- Haar, 32
- Hankel, matrice de, 87
- harmoniques sphériques, 109
- harmoniques, nombres, 180
- Hermite, 151
- Hermite, polynômes d', 93, 102, 107, 112
- Hermite-Genocchi, 155
- hermitien (form.), 101
- Hilbert, espace de, 130

- intégration, Gauss-Legendre, 97
- intégration, Gauss-Tchebycheff, 99
- intégration, formule de Gauss d', 96
- Intégration: Clenshaw-Curtis, 161
- Intégration: plusieurs var. et soft. KULeuven, 161
- Intégration: règle de Clenshaw-Curtis, 161
- Intégration: règles adaptatives, 161
- interpolation d'Hermite, 146

- Interpolation rationnelle, 149
 irrationnalité, 16
 Ismail, M.E.H., 107

 Jackson, th. de, 134
 Jacobi, polynômes de, 102, 107
 Jordan, théorème de, 56

 Karatsuba, 126
 Kolmogorov, 27, 34
 Kronrod, 161

 Lagrange, 145
 Lagrange, base de, 147
 Laguerre, polynômes de, 93, 102, 107
 Laplace-Beltrami, 109
 Lebesgue, constante de, 58
 Lebesgue, lemme de, 58
 Legendre, polynômes de, 91, 92, 107, 110
 Leibniz, formule de, 152
 Lissajous, figure de, 39
 Lobatto, 105, 161

 Müntz, 82, 137
 Markoff, 42
 Mawhin, J., 17, 104, 135
 médiane, 26
 Meinguet, J., 8, 19, 45, 148, 149, 173
 module de continuité, 62
 moindres carrés, 114
 moments, problème des, 140
 moyenne, 26

 Nevai, P., 107
 Neville-Aitken, 151
 Newton-Cotes, formules de, 170
 norme, 20, 24
 norme 1, 118
 norme stricte, 24
 noyaux, 137
 NURBS, 142

 ondelettes, 126
 orthogonal, orthogonaux, 74, 79
 orthogonale, matrice, 95
 orthogonalité: pol. de Tchebycheff, 51
 orthogonalité: Sturm-Liouville, 51

 Padé, 67
 Parseval, 121, 130
 Peano, théorème de, 162
 Penrose, 80
 Plana, formule sommatoire de, 182
 Poisson, formule sommatoire de, 183
 Poncelet, 14
 prehilbertien, 27
 prehilbertien, espace, 76
 premiers, nombres, 70
 produit scalaire, 73

 qd, algorithme, 91
 quasi-périodique, 80
 quotient-différence, algorithme, 91

 Radau, 161
 rationnelle, approximation, 40, 45, 67
 rationnelle, interpolation, 149
 récurrence, 39, 44, 45, 88, 92
 relative, erreur, 32, 40
 Remez, 33, 43
 représentant de Riesz, 137
 Riccati, 17, 105
 Riemann-Lebesgue, 56, 121
 Rivoal, 18
 Rodrigues, 51, 104, 105
 Romberg, schéma de, 179
 Ronveaux, A., 104
 Runge, 158

 SIAM, 10
 Simon, B., 107
 Simpson, formule d'intégration de, 170, 171, 173
 sommatoires, formules, 180, 182
 spline, 26, 156
 Stirling, formule de, 182
 Stone-Weierstrass, 135
 Sturm-Liouville, 48, 51, 101

 tau, méthode des, 65
 Tchebycheff, 68
 Tchebycheff, centre et rayon, 27
 Tchebycheff, polynômes de, 35, 41, 107
 Tchebycheff, séries de, 57
 Tchebycheff, théorème de, 28
 totale, suite, 129
 trapezes, règle des, 170, 171
 Trefethen, L.N., 68

 unicité, 23, 24, 26, 30
 unicité, forte, 25, 31
 unitaire, matrice, 95

 Van Pée, M., 173
 variation bornée, 56, 121

 Weierstrass, théorème d'approximation de, 62, 131

 Zéros de polynômes, 93
 zeta, fonction- de Riemann, 18, 70, 182
 Zolotarev, 40
That's all, Folks